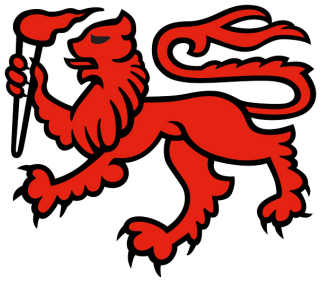


UNIVERSITY*of*
TASMANIA

Temperature reconstruction methods

J. D. Vlok

School of Engineering
College of Sciences and Engineering

29 April 2019

Last update: 15 May 2019

http://eprints.utas.edu.au/29788/

jdvlok@utas.edu.au

Executive Summary

The surface air temperature of Earth is historically measured at a collection of individual weather stations. Deriving regional averages and long-term trends from these point sources require sophisticated mathematical analysis and algorithm development. To understand historical temperature, non-climatic artefacts written into the record must first be identified and removed. These artefacts include everything not attributable to the true climate, including changes in the environment surrounding weather stations, and changes in measuring methods and instrumentation. Data quality also varies and in some cases temperature measurements are missing or incorrect.

The historical temperature record must hence be reconstructed from a collection of individual time series. Reconstruction involves quality control to remove suspicious data, infilling to recover missing data, and spatial interpolation to estimate temperature series for locations where no measurements were ever taken. From the reconstructed record, averages and trends can be determined for individual locations and larger regions.

This report describes temperature measurement and reconstruction with a focus on Australia over 1910 to 2018. The official homogenised reconstruction of the Bureau of Meteorology is considered, and two alternative methods are presented with some comparisons. The alternative methods include nearest-neighbour infilling and an approach based on artificial neural networks. An overview of existing spatial interpolation techniques is also given, aiming to identify suitable benchmarks for evaluating alternative reconstruction techniques.

The nearest-neighbour technique was used to infill all raw monthly mean temperature data available for Australia. Average anomalies were calculated from the infilled record, which show a good match with the official reconstruction of the Bureau, especially for an area where stations are distributed approximately uniformly.

The neural network method is illustrated conceptually in this report, with results indicating improved performance compared with existing benchmark methods, when using the technique to estimate temperature data for unsampled locations. Further development of the neural network method is required to improve performance and to estimate trends and averages of larger regions.

Contents

List of Figures

List of Tables

Acronyms

ACORN-SAT	Australian Climate Observations Reference Network - Surface Air Temperature
ADAM	Australian data archive for meteorology
AI	artificial intelligence
ANN	artificial neural network
AWS	automatic weather station
BoM	Bureau of Meteorology
CDO	climate data online
CI	confidence interval
CO_2	carbon dioxide
GISS	Goddard Institute for Space Studies
GSL	Gnu Scientific Library
IDS	inverse distance squared
IDW	inverse distance weighting
IEW	inverse exponentially weighting
IPCC	International Panel on Climate Change
LCL	lower control limit
LiG	liquid-in-glass
LOOCV	leave-one-out cross-validation
MA	moving average
MAE	mean absolute error
ML	machine learning
MR	moving range
MSE	mean square error
NASA	National Aeronautics and Space Administration
NCEI	National Centers for Environmental Information

NOAA	National Oceanic and Atmospheric Administration
OCR	optical character recognition
PDF	probability density function
PM	percentile-matching
PRT	platinum resistance thermometer
QC	quality control
RTD	resistance temperature detector
RV	random variable
SAT	surface air temperature
SGD	stochastic gradient descent
SPC	statistical process control
TIN	triangular irregular network
TPS	thin plate splines
TSA	trend surface analysis
UCL	upper control limit
UHI	urban heat island

1 Introduction

1.1 Analysing historical temperature

It is believed that Galileo constructed the first thermoscope, a device to detect air temperature change, in the early 1590s. About 130 years later, Fahrenheit invented the mercury thermometer with a standardised scale, although a theoretical understanding of temperature was still undeveloped [1].

The widespread measurement of the earth's surface air temperature (SAT) using networks of weather stations commenced in the 1850s, although the single longest record dates back to 1659 [2]. In most countries, weather observation networks were developed mainly to support weather forecasting. Monitoring long-term climate change has only become a priority since the early 1990s [3], after which several research groups have created datasets to describe variations in the global average temperature over time.

Such historical reconstructions include the NASA GISS[1], HadCRUT[2], NOAA NCEI[3], and Berkeley Earth[4] datasets. The NASA annual global mean SAT reconstruction over 1880–2018 is shown as an example in Fig. 1 below, with data obtained from [8].

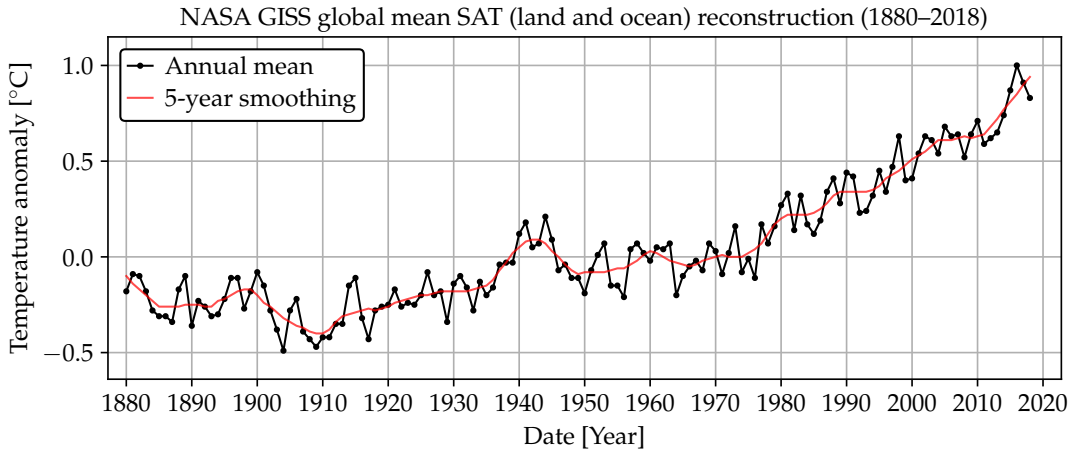


Figure 1: Global mean temperature series with reference period 1951–1980, from [8].

[1] The National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies (GISS) temperature series starts in 1880 [4].

[2] A dataset formed by the UK Met Office (Hadley Centre) and the University of East Anglia (Climatic Research Unit) starting in 1850 [5].

[3] The US National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI) dataset starts in 1880 [6].

[4] An independent non-profit organisation that compiled a dataset dating back to 1753 [7].

To understand past climate before the instrumental record, several studies have been conducted using proxy data from sources such as tree rings, fossil pollen, ice cores, ocean sediments, corals, etc. [9]. Fig. 2 shows a conceptual diagram of warm and cold periods reconstructed from proxy data and the instrumental record, based on timelines and graphs provided in e.g. [10–12]. The instrumental record is indicated by the relatively warm current warm period (CWP) after 1850 in Fig. 2.

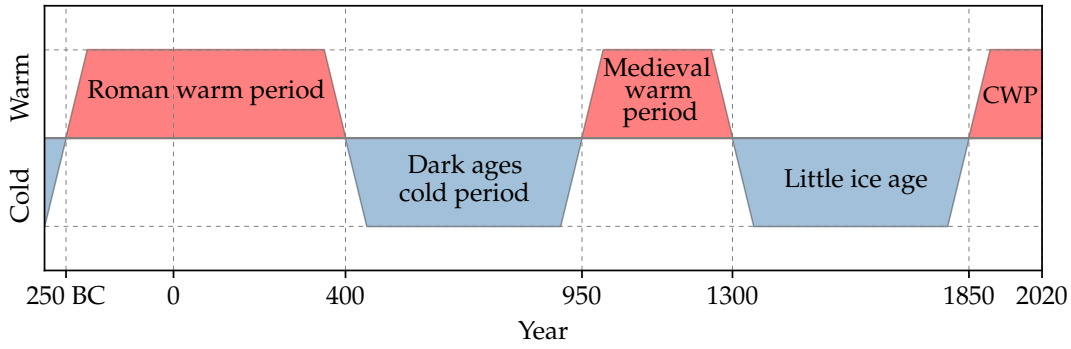


Figure 2: Schematic diagram of historical temperature variation.

The historical reconstruction illustrated in Fig. 2 can only be considered global after 1850 (generally with increased accuracy and quality as time progresses), as temperature has only been recorded widely over the earth's surface from this date. To calculate a true global average, data from sensors worldwide must be processed using sophisticated mathematical techniques, including spatial estimation and interpolation methods.

To reconstruct a global picture before 1850 is more challenging as disparate proxy data must be combined. Although several studies focused on different isolated locations indicate similar historical temperature patterns as depicted in Fig. 2, the local or global extent of climate before the instrumental record remains less clear (see for example [13]).

1.2 Potential drivers of temperature

James Watt patented the steam engine in 1769, marking the start of the industrial revolution and a sudden and enduring increase in atmospheric carbon dioxide (CO_2) levels [14], as indicated in Fig. 3. Although the natural flow of CO_2 into the atmosphere is larger than the contribution due to humans burning fossil fuels, the natural flow of CO_2 out of the atmosphere into the biosphere and the oceans is believed to approximately balance the natural inflow. The conclusion is then that rising atmospheric CO_2 levels are mainly caused by human activity [14].

By comparing the instrumental global mean temperature record (Fig. 1) and the associated atmospheric CO_2 level (Fig. 3), the correlation is undeniable – both temperature and CO_2 increase together.[5] Although correlation does not necessarily imply causation [15], the rising atmospheric CO_2 level has widely been accepted as the principal driver of increasing temperatures. This argument is supported by studies showing that rising temperature is preceded by increasing CO_2 levels (see e.g. [16]), and the physical process explanation offered by the greenhouse effect [17, 18]. Alternative explanations have however also been put forward, where changes in temperature and CO_2 are hypothesised to be part of a natural process, largely independent of human industrial activity [19–21].

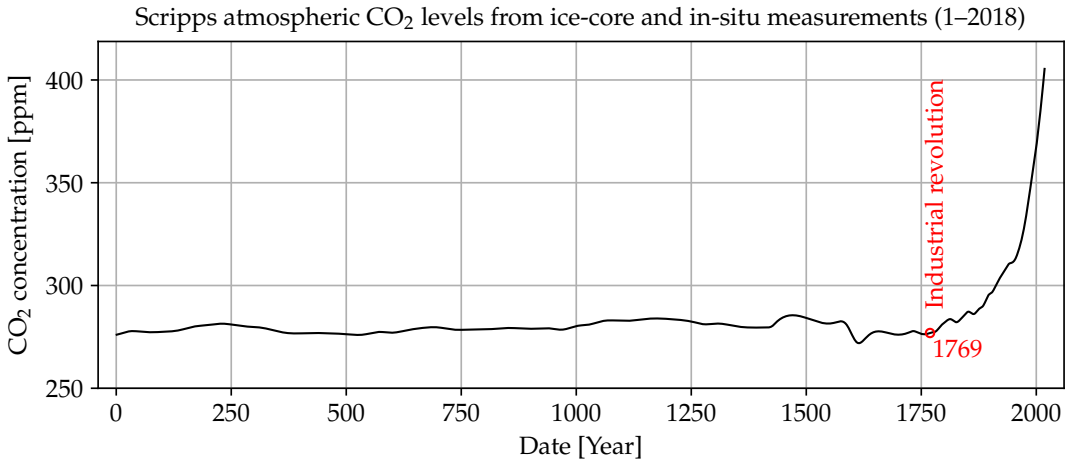


Figure 3: Atmospheric carbon dioxide reconstruction, from [22–24].

Whereas the International Panel on Climate Change (IPCC) predicts that global warming will likely reach 1.5 °C above pre-industrial levels between 2030 and 2052 [25], one study predicts a solar grand minimum will occur over this same period (solar cycles 26–27) due to the behaviour of the sun's magnetic fields [26], resulting in much cooler temperatures. As indicated in Fig. 4, reduced solar activity (e.g. the Maunder minimum spanning 1645–1715) is historically an indicator of cooler periods (e.g. the Little Ice Age illustrated in Fig. 2).

Volcanism offers an alternative (or parallel) explanation of historically cooler periods to a reduction in solar activity. A Berkeley Earth study argues cooler periods can be associated with volcanic sulfate emissions, without considering fluctuations in solar activity (where the period after 1750 was studied). Including solar activity in addition to CO_2 and volcanic eruptions to describe the earth's historical temperature record did not significantly improve their model [7].

[5] A close match can be obtained by applying mathematical transformations, e.g. the natural logarithm of the CO_2 curve approximately matches the Berkeley Earth temperature reconstruction [7].

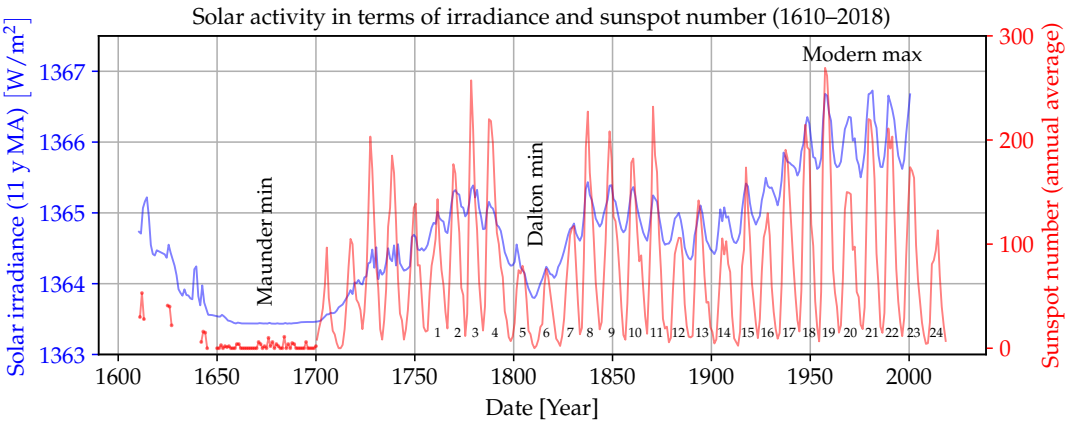


Figure 4: Solar activity including irradiance [27] and sunspot number [28, 29].

There are many other drivers and factors influencing climate in varying degrees over time that form an intricate feedback system, including other greenhouse gases such as water vapor [30] and methane [31], the behaviour of other bodies in the solar system including Jupiter [32] and the moon [33], ocean currents, plate tectonics, changes in the cover and usage of the earth's surface, the earth's tilt and wobble, etc. [34].

Central to the debate of climate change and the main drivers thereof, is the historical temperature record. Understanding temperature measurement and reconstruction methods is essential before relationships with other environmental variables can be analysed.

1.3 Reconstructing historical temperature records

The reconstruction of historical temperature involves creating new data records from available temperature observations, and/or other observations where a relevant relationship with temperature is known. The new record should be more complete or improved, or should describe temperature characteristics that are not readily apparent or accessible in the original data.

Reconstruction may involve creating continuous records for single locations, larger areas or regions, entire countries or continents, oceans, or the entire earth. An example is estimating global average temperature over time using several individual records as mentioned in Section 1.1 above. Reconstructing temperature data may also involve the following processes.

- Recovering missing data within individual records
- Estimating data values for unknown locations

- Performing quality control (QC) and correcting erroneous data values
- Establishing or uncovering relationships between temperature and other variables, which may be utilised in performing the above tasks
- Calculating average values or other statistics, such as confidence levels of estimates

Apart from the potential climate drivers mentioned in Section 1.2 above (including atmospheric CO_2, solar activity and volcanic sulfate emissions), local temperature is largely determined by the following factors.

- Location (latitude, longitude, altitude), which is essential in estimating missing data and performing spatial interpolation (see Section 4). Temperature generally decreases with increasing distance from the equator and with increasing altitude[6].
- Morphology of the area and surrounding landscape, which define how regional climatic elements (e.g. wind, cloud cover, rain) affect the location.
- Regional factors, e.g. land use and proximity to water bodies, arid land, forest, etc.
- Atmospheric conditions, including presence of aerosols such as dust, smoke, water vapor, etc.

Location data are typically readily available, often packaged with temperature data. Data related to the other above-mentioned factors are not that easily obtainable and often change over time. For example, the distance to water bodies is subject to change as rivers flood and dry. Local land use and morphology can also change with construction and developments.

Some of these data elements can however be extracted from e.g. geographical maps and historical news archives and may then be utilised in temperature reconstruction. Nevertheless, such data are also inherently encapsulated within temperature observations, and can be used indirectly if the correct reconstruction approach is followed. These themes are further considered in Sections 2 and 3.

One motivation for reconstructing historical temperature is to understand past climate and the relationships with other environmental variables, in order to predict what future temperature (and climate in general) might look like.

The focus of this report is to reconstruct historical temperature from the raw instrumental record of Australia using a number of interpolation techniques and to make comparisons with official results.

[6] The rate at which temperature changes with altitude is known as the *lapse rate*.

1.4 Instrumental temperature record of Australia

Daily manual measurement of minimum and maximum temperatures across Australia commenced in the 1840s with the number of weather stations gradually increasing thereafter. Since 1910 the Australian Bureau of Meteorology (BoM) established standardised equipment to measure SAT through the widespread employment of the Stevenson screen, and the application of certain specifications to ensure measurements are taken in similar conditions across the continent [35].

There are nearly 2000 locations across Australia where temperatures have been measured and recorded in the Australian data archive for meteorology (ADAM), publicly available through the BoM climate data online (CDO) portal. The locations are shown in Fig. 5, although the measurement network and environment constantly changed over history. Many stations were closed or relocated, new stations were opened, buildings were developed around stations, and measurement technology changed. Furthermore, stations are roughly distributed according to population density, creating a long data history around south eastern Australia, but leaving a relatively sparse record for remote areas.

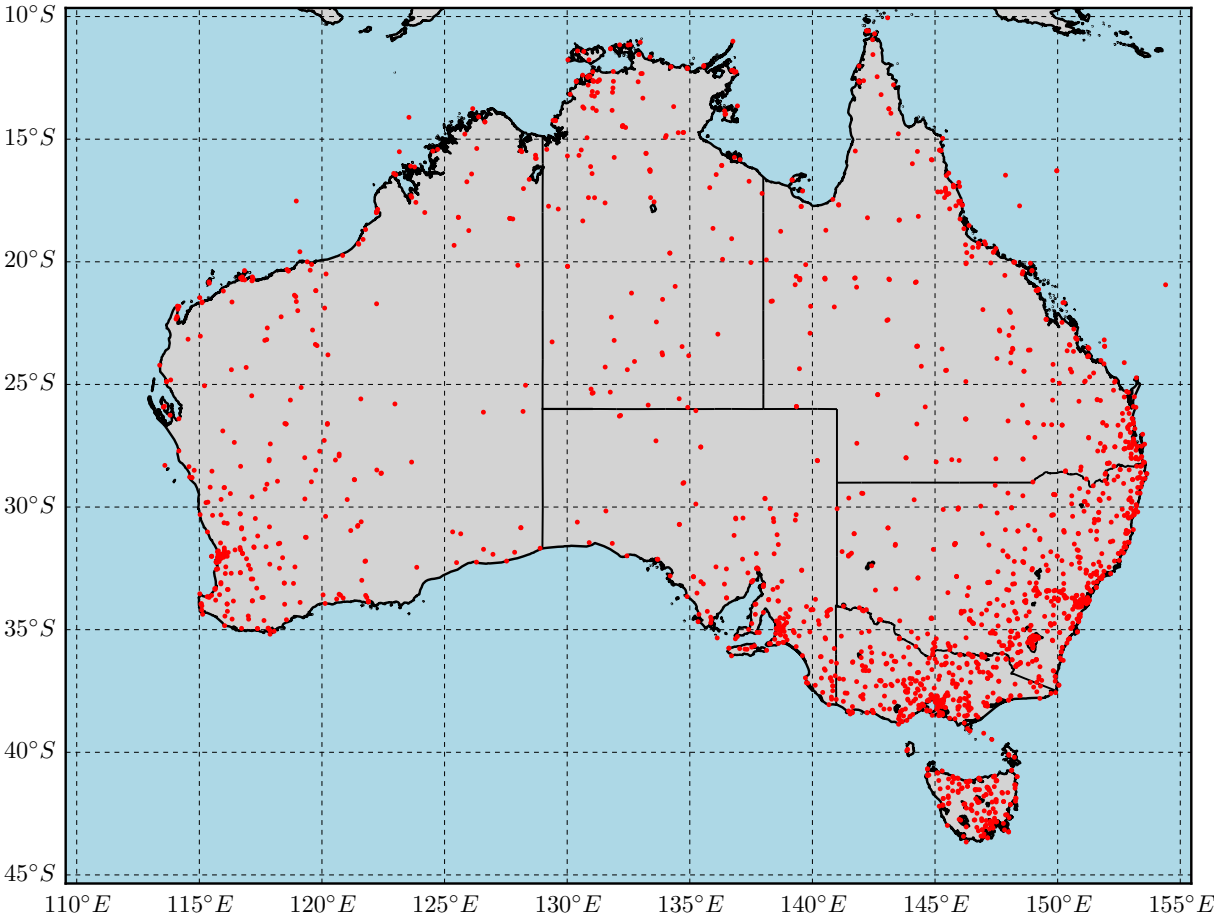


Figure 5: Locations of all SAT-measuring stations across Australia between 1840 and 2018.

There are 1895 stations shown in Fig. 5 that contributed monthly mean maximum ($T_{\max}$) data to the ADAM database[7]. The individual station records are illustrated as twelve-month moving averages (MAs) in Fig. 6a, with the average in black. The number of measurements for any given month is shown in Fig. 6b. Considering all stations shown in Fig. 5 over 1841–2018, the record shown in Fig. 6a can be said to be only 18.9% complete. If the early record is not considered, this value improves to 29.2% over 1910–2018.

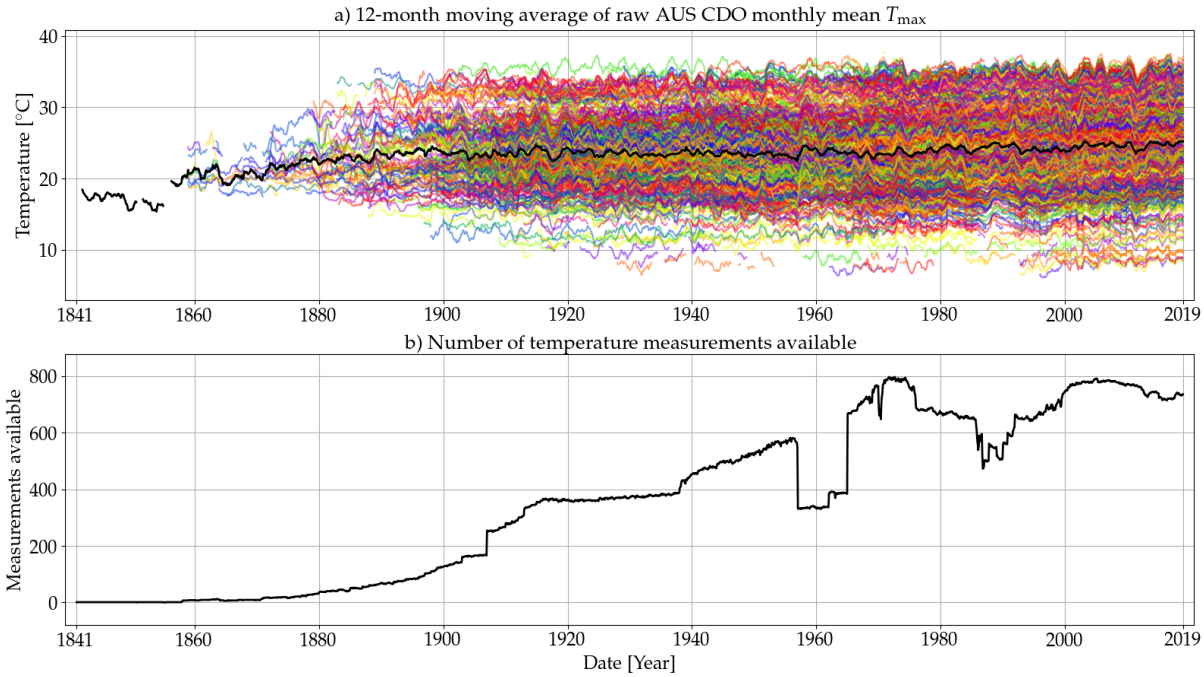


Figure 6: Depiction of all monthly mean maximum temperature data (1841–2018).

The black line in Fig. 6a depicts the station mean over time, clearly containing bias due to the varying number of active stations. The bias also inherently contains a geographical component as the mix of stations, representing a set of locations, varies over time. The station mean itself is therefore not an accurate representation of long-term variability, and an improved estimation of the average trend can be obtained by removing the geographical and temporal bias, which may be achieved by

- infilling or interpolating all missing data of individual records, such that all records are complete over the desired period (e.g. 1910–2018), and
- estimating temperature data for unsampled locations

which form part of the reconstruction process as mentioned in Section 1.3. The temperature history of locations where no weather station ever existed can then also be described, and a geographically-representative or area-weighted average could be calculated over time to derive long-term trends.

[7] Including all the data available via the BoM CDO portal according to the weather station directory [36].

2 Changes in the sensor network and environment

The earliest monthly mean temperature record available in ADAM is that of the Hobart Botanical Gardens (ID 94030). The $T_{\max}$ record for this station runs from January 1841 to December 1854 (with a few additional values in 1880), and is the only series shown for this period in Fig. 6.

Around 1860 the number of weather stations started to increase gradually, although it was only in 1910 that temperature measurement was standardised - shortly after the BoM started operating as a Commonwealth agency on 1 January 1908. Before 1910, SAT was measured using various configurations, including thermometers housed in beer crates on outback verandas [35].

The aim of standardisation of equipment is to maximise consistency over space and time in monitoring weather variables. Measures taken by the BoM include housing thermometers in a protective enclosure known as a Stevenson screen, 1.2 m above ground level on a natural surface separated from nearby structures such as buildings or plant growth. Also, measurements are taken traditionally at 9h00 and 15h00 every day and recorded in field logbooks by trained observers [37].

Although standards and specifications were put in place to minimise the variance in weather observations, technological advances and expansion of the sensor network, coupled with changes in the environment surrounding weather stations introduced new challenges with non-climatic artefacts encapsulated in the record. Some of these contributing factors are subsequently discussed.

2.1 From liquid-in-glass to electronic thermometers

The classical instrument for temperature measurement is the liquid-in-glass (LiG) thermometer, including mercury-in-glass for measuring maximum temperature and alcohol-in-glass for minimum temperature[8]. Daily minimum and maximum temperatures can be measured using registering LiG thermometers where an index marker is moved by the liquid in the glass tube to capture the daily extremes, before being manually reset [38].

Whereas LiG thermometers rely on the principle of thermal expansion of liquids to mea-

[8] The temperature range of operation between the melting/freezing and boiling points of mercury is approximately -39 to 357 °C, and of ethanol -114 to 78 °C.

sure temperature, greater accuracy can be achieved by exploiting the relationship between temperature and the resistance of electrical conductors. The resistance temperature detector (RTD) has become a popular alternative in measuring temperature, and also offers a wider range of measurement and improved safety.

The platinum resistance thermometer (PRT) is one type of RTD and has a very stable and linear resistance-temperature characteristic over a wide range of temperatures (-200 to 650 °C, although some types can reach 800 °C) for various industrial applications [39]. The BoM introduced custom-designed[9] PRTs into the Australian temperature sensor network in 1992, commencing the era of automatic measurement and logging of SAT [40].

Since 1992 temperature has therefore been measured using two technologies - the classical LiG thermometers and the newer electronic PRT probes. As shown in Fig. 7, the number of stations using LiG thermometers was reduced to around 200 at the end of 2018, while automatic weather stations (AWSs) - which utilise PRTs - increased steadily[10].

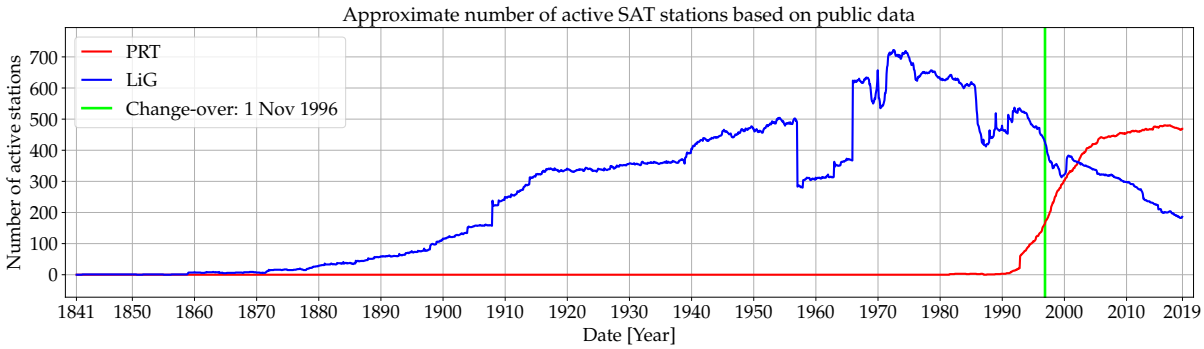


Figure 7: Evolution of changing temperature sensors across Australia.

There are however a number of weather stations still employing both technologies. From 1 November 1996, the PRT became the official measuring instrument, and human operators were no longer allowed to overwrite PRT values with LiG measurements [40], indicating that there may be differences between the two sensors. Furthermore, an initial analysis into a limited set of parallel data suggests that the two sensors are not producing statistically equivalent data, which may be ascribed to a difference in time constants and/or the method used to average measurements [41].

[9] The BoM designed the platinum probes to approximate the time constant of mercury thermometers within ± 5 seconds, by changing the dimensions of the protective metal sheath [40].

[10] At the end of 2017 there were 563 AWSs measuring SAT and 212 stations still utilising only LiG thermometers (BoM, pers. comm. 2017). The numbers differ from those shown in Fig. 7 as public data were used to create the graphs in the figure - some stations may have been inactive, or not publicly accessible.

Note that the ADAM database does not include any parallel recordings - a single time series is available for every weather station consisting of LiG measurements up to the date the sensor was replaced by a PRT, followed by measurements obtained only by this new electronic sensor. These transition dates are available collectively at [42, 43] (last updated June 2012) and individually in the online "basic site summary" climatological station metadata [44].

2.2 Urban heat islands and land modification

The urban heat island (UHI) effect is a phenomenon where urban SAT is higher than surrounding non-urban regions due to differences in areal thermodynamics, caused by urbanisation including an increase in building developments, concrete and asphalt roads. Fig. 8 displays temperature data for an urban area (Melbourne Regional Office, which was located in the city) in comparison with a regional area (Echuca airport) separated by 200 km. Although the two locations had similar minimum temperatures between 1880 and 1950, a clear departure in this pattern started around 1950. Note that the two maximum temperature series do not exhibit this bifurcation.

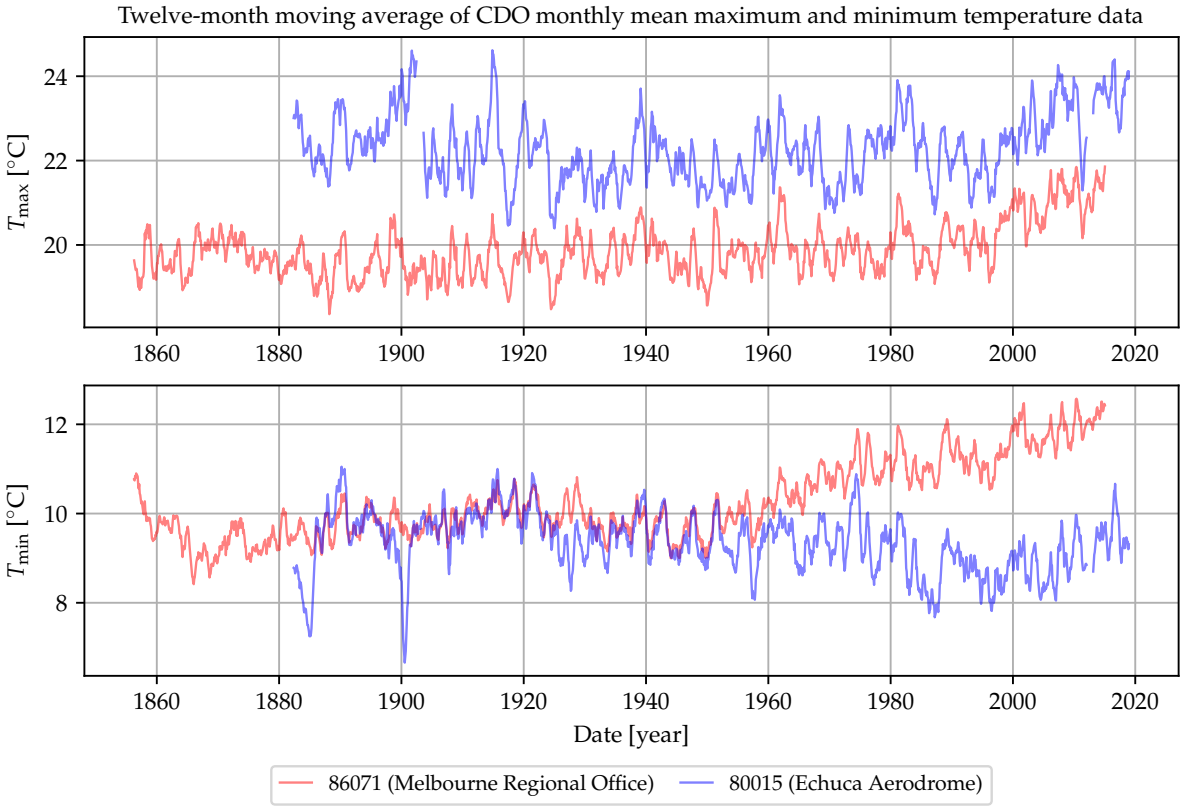


Figure 8: Maximum and minimum temperature series to illustrate the UHI effect.

This pattern could be explained by the fact that buildings and asphalt heat up during the day and radiate the heat back when the surrounding air cools down, effectively raising the minimum temperature.

To analyse trends in more detail, knowledge regarding changes in land use over time (e.g. building development and progress timelines) should also be utilised, although such information is not always documented. Furthermore, any modification of land surface may affect the temperature. For example, vegetation growth and agricultural development such as construction of irrigation systems, rivers and dams will most likely change the regional temperature profile [45].

Temperature data originating from UHIs (or any area with significant change in land use) will adversely affect estimates of global climatic change. One study found that non-climatic factors including land surface modification contributed around 50% to the global warming trend over 1980 to 2002 [46]. To improve estimates of long-term temperature change, a number of approaches to isolate and remove or correct UHI-contaminated data have been suggested, including analysing weather station metadata, thermal infrared images and nightlight data from satellite images [47].

2.3 Other effects

To accurately monitor air temperature of a given location over an extended period of time, there should ideally be no change in the sensor, the weather station configuration or the surrounding area. Any change in measurement method or the surroundings will introduce non-climatic artefacts into the record and limit the ability of the sensor to detect changes in the climate, and to draw conclusions regarding long-term changes in temperature.

Apart from changes in measurement technology and the UHI effect discussed above, the following factors may also introduce non-climate artefacts into the temperature record.

1. **Changes in weather station location**
   Moving a sensor to a new location is similar to changing the area surrounding a static sensor as discussed in Section 2.2, as in both cases the environment being sampled changes. A common trend worldwide is the relocation of weather stations from town centres to often cooler airports, which should induce an artificial cooling trend if not corrected [48].

2. **Isolated local effects**
   Incidents such as wildfires, affecting only the local temperature around one or more sites may not necessarily reflect the regional atmospheric conditions accurately. Depending on the application of weather data, these local measurement may need to be excluded, e.g. when calculating regional temperature trends.

3. **Changes and maintenance of weather station equipment**
   Apart from upgrading LiG thermometers to electronic probes (see Section 2.1), weather station equipment needs to be maintained. For example, replacing a broken thermometer or repainting a Stevenson screen may affect the historical record.

4. **Use of different types of Stevenson screens**
   Although the BoM standardised their weather stations in 1910, different Stevenson screen sizes have been used, which may impact temperature measurement [49]. However, records of screen type in use over history are available from the BoM [40].

5. **Measurement accuracy**
   Temperatures are currently measured to 0.1 °C accuracy in Australia. In the past, observations were rounded differently and °F was more popular. Furthermore, the process of measuring, reading and transmitting temperature data is subject to error at each phase, caused by a combination of faulty equipment and human error [50].

6. **Time of observation**
   In Australia temperatures were traditionally recorded at 9h00 and 15h00 local time at most stations, with a difference of periods over which minimum and maximum daily values are determined [3]. Time zones and daylight saving may therefore impact measurements. Furthermore, the newer AWS allows near-continuous measurement and reporting of the weather.

7. **Algorithms used to calculate average values**
   The mean daily temperature was defined traditionally as the average of the minimum and maximum daily values. However, with advances in technology, mean daily temperature can be estimated more accurately as measurements can easily be taken more frequently [48]. Also, changes in the calculation of mean monthly temperatures can substantially influence trends [51].

Some factors influence single stations in isolation at various times, while other factors influence many stations collectively, such as when observation practices are changed over large areas. The possible impact of all these factors should be considered when analysing historical climate data, especially when deriving long-term trends.

3 Homogenisation and the official reconstruction

All irregularities or inhomogeneities contained in climate records should be removed, before using these records to describe the climate history[11]. These inhomogeneities include all contributions not describing climate, caused by changes in e.g. weather station location, the environment surrounding each weather station, measurement equipment, and observation practises [52].

This section will consider some methods that are used to detect and remove such non-climatic artefacts present in the historical temperature record. The official historical temperature reconstruction developed by the Australian BoM will also be discussed.

3.1 Homogenisation overview

The process of detecting and removing non-climatic effects from climate data is known as homogenisation. The aim is to correct artificial changes (typically by adjusting raw measured data) to create a homogeneous record, which is presumably more consistent over time and that more accurately reflects the true climate history. Such a record can then be used to draw conclusions regarding long-term climate patterns.

The process of homogenising climate data relies on one or more of the following [53].

1. Statistical and graphical methods to identify and correct abrupt changes or break points in data series [54].
2. Documentary records or station history metadata, which contain dates and other information regarding changes to a weather station and its equipment.
3. Parallel data records, or similar weather measurements taken at neighbouring weather stations over the same time period. The methods mentioned in point 1 are typically used on these parallel records in combination with the data to be homogenised from the candidate station.

Several homogenisation techniques have been developed and differ in how much emphasis is given to each of the above components. Two further classifications are considered below [53].

[11] Some factors, e.g. the UHI effect, are undesirable when studying long-term atmospheric temperature trends, whereas the same factors could be important for other studies, such as evaluating the temperature trends in developing cities.

3.1.1 Objective and subjective homogenisation

Objective homogenisation techniques detect changes and adjust the data automatically according to some algorithm. The advantage of these techniques is their reproducibility and ease of processing large datasets [53]. However, there may be different ways of implementing automatic techniques in software and human intervention may still be required in a small subset of the data (there may be a few border cases that need special attention). Attaining full objectiveness and exact reproducibility may therefore be difficult.

Subjective homogenisation techniques mostly rely on judgments made by climate experts [53]. Subjective assessment is especially needed when dealing with incomplete historical measurements and metadata where large uncertainty is present, e.g. where records are inconclusive or contradictory, or where a variety of sources is used (e.g. newspaper archives documenting changes in weather station locations). Using subjective methods could further be justified based on the unique circumstances and history of each individual weather station [35, 55, 56].

Homogenisation techniques can also consist of both objective (manual) and subjective (automatic) elements, creating semi-automatic methods.

3.1.2 Absolute and relative methods

Absolute homogenisation methods only consider individual station records in isolation, including measurements and metadata for each station. These methods are therefore limited to applying statistical tests on single time series, and cannot distinguish between natural and artificial causes of discontinuities, except if supported by station metadata. The capability of detecting true climate signals from single station records is therefore also limited [57].

Relative homogenisation methods compare data from the candidate station (to be homogenised) with neighbouring or reference stations. For example, the difference time series between the candidate and reference stations can be used to detect inhomogeneities, assuming nearby stations are sufficiently synchronous (as they are exposed to approximately the same climate signal). The performance of relative methods can be improved by selecting reference stations according to some criteria, such as ensuring that each reference station is homogeneous over a specified time period and that it is highly correlated with the candidate station [58].

3.2 The official Australian reconstruction

The BoM created a long-term temperature record known as the Australian Climate Observations Reference Network - Surface Air Temperature (ACORN-SAT) to monitor climate variability and change in Australia [35]. The dataset contains homogenised daily records for 112 locations on the main Australian continent (including Tasmania) shown on the map in Fig. 9, and monthly records for 8 remote locations (remote islands and Antarctica) [59]. ACORN-SAT only includes data after and including 1910, as climate observations prior to 1910 were limited and standard observation methods were not yet specified [60].

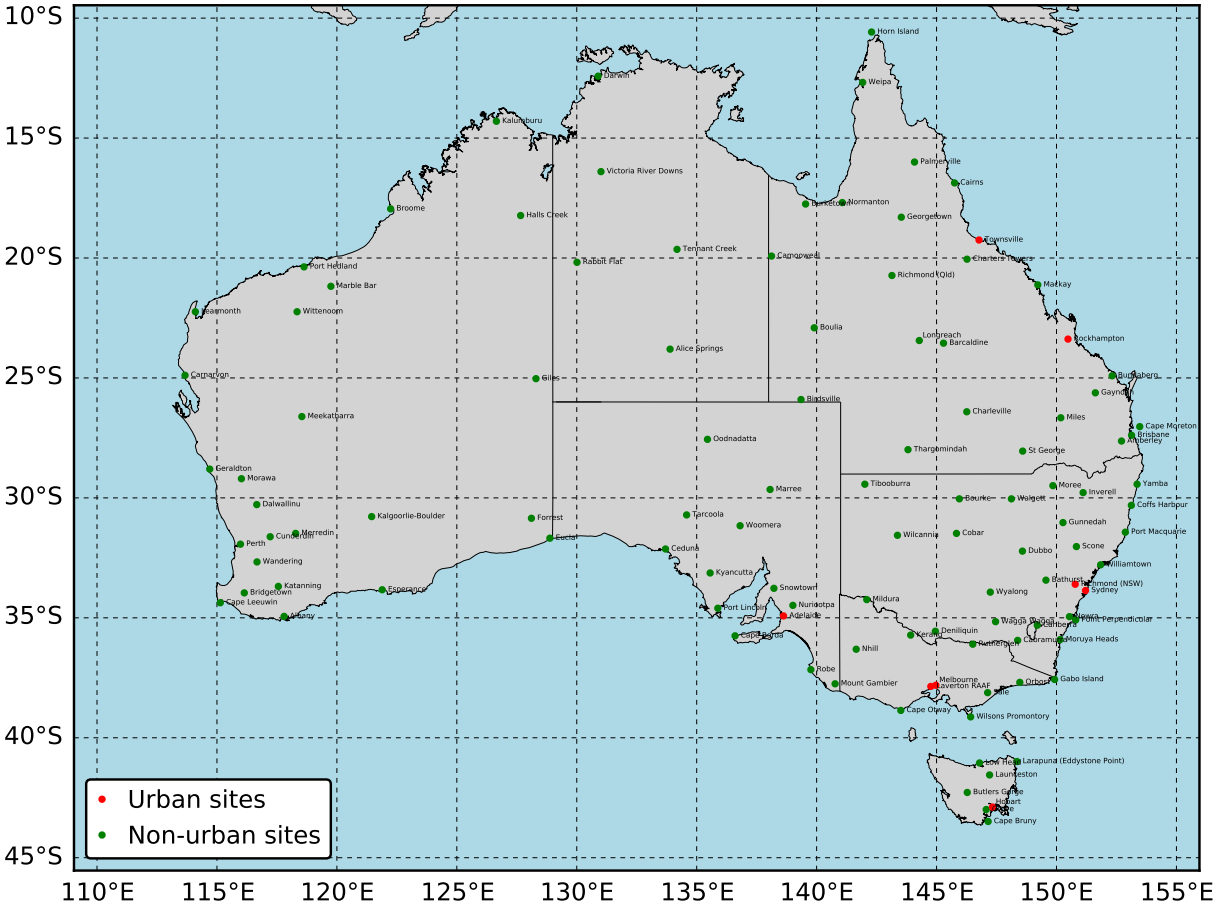


Figure 9: Locations of 112 ACORN-SAT sites across Australia.

The ACORN-SAT dataset incorporates temperature records obtained from weather stations located at the 112 sites shown on the map, and an additional 86 records from nearby stations [61]. Also, the 112 locations include 8 urban sites as indicated, which are excluded in the calculation of regional averages to reduce the UHI effect.

3.2.1 Homogenisation of individual series

To create individual homogenised series, the BoM uses statistical methods in combination with metadata. Relative methods are applied when sufficient neighbouring data records exists, which is the case for most of the 112 locations shown in Fig. 9. For the remote islands and Antarctica dataset, nearby reference stations typically do not exist and homogenisation relies mostly on performing statistical tests on single series in consultation with metadata [60].

The ACORN-SAT homogenisation process for stations with sufficient neighbours consists of the two stages summarised below, after removing daily measurements that failed QC checking [3].

1. **Detection of inhomogeneities**
   Potential break points in the candidate time series are identified through pairwise comparison with a number of reference stations using different time scales [62]. Reference stations are chosen by first identifying the nearest 150 neighbours, and then selecting 40 stations from this pool that have the highest correlation with the candidate station. The candidate station must also have at least 7 reference sites for each year, otherwise stations with smaller correlation values are used instead of the top 40. Finally, the time series containing the inhomogeneity (either the candidate or reference series) is identified and consolidated with metadata.

2. **Removing inhomogeneities through data adjustment**
   After breakpoints have been identified, the data is adjusted and/or merged by applying a non-linear transfer function to correct the frequency distribution of daily values caused by the breakpoint, as it was found that inhomogeneities typically affect temperature values non-linearly. For example, lower temperature values could change more than higher values after a site move, requiring a non-linear correction [35]. The transfer function is calculated using one of the percentile-matching (PM) algorithm variations [52], depending on the amount of overlap between neighbouring stations.

Fig. 10 shows the homogenised ACORN-SAT series for Rutherglen, including daily T_{min} obtained from [63] and daily T_{max} obtained from [64]. Each series starts on 8 November 1912 and ends on 28 February 2019, with all values over 1960–1964 missing. Counting from 1910, nearly 10% of each series is therefore missing. In total for the 112 sites shown in Fig. 9, approximately 18% of the data are missing over the period of standardised measurements.

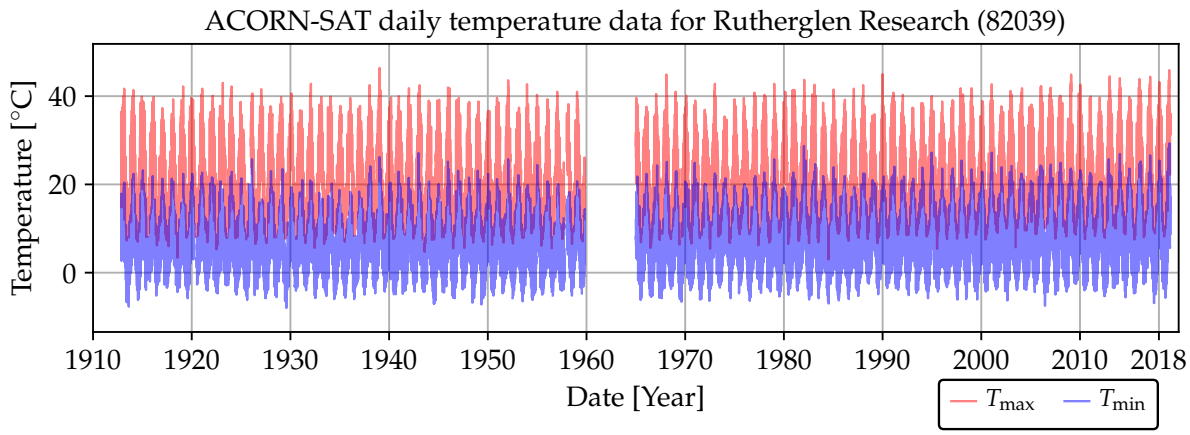


Figure 10: ACORN-SAT daily minimum and maximum series for Rutherglen.

3.2.2 Calculation of regional averages and trends

To calculate average temperatures for large regions (e.g. Australia as a whole, entire states or other areas of interest such as the Murray Darling Basin), a regular grid with 25 km resolution (approximately $0.25° \times 0.25°$) is used as area-weighting mechanism. The contribution of each temperature measurement in the average is thus proportionally scaled according to each station footprint or land area being represented. The calculation of average temperatures is performed as follows [65].

1. Monthly averages are calculated using the homogenised daily series for each ACORN-SAT site, allowing a maximum of 10 days missing per month.

2. The monthly normal or reference value, which is the average over the 1961–1990 reference window, is calculated for each station.

3. The temperature anomaly series is calculated for each station, which is the difference between monthly values and the reference value, indicating the departure (or anomaly) from the reference period for each month.

4. The station anomalies are interpolated to the regular grid using the Barnes successive correction algorithm [66], while excluding UHI-contaminated stations shown in Fig. 9.

5. Regional average anomalies are calculated from these interpolated grid-point values, which are effectively area-weighted values obtained from the homogenised data.

Some results calculated by the BoM using the above process are displayed in Fig. 11, including the T_{max}, T_{min} and T_{mean} anomaly series for the Australian continent [67].

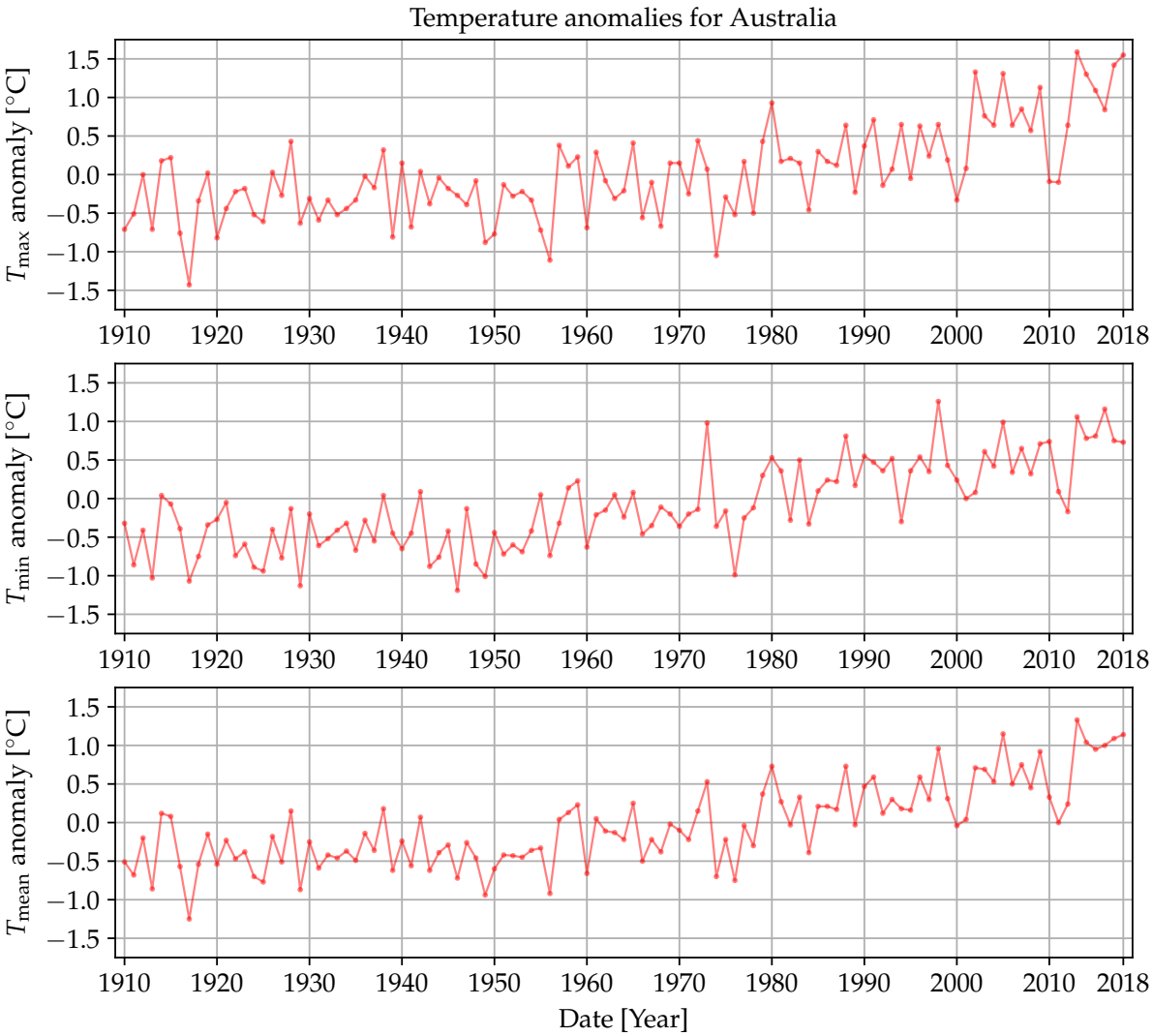


Figure 11: Official temperature anomaly series for Australia.

3.2.3 Historical development of Australian reconstructions

Three major homogenised temperature datasets have historically been developed for Australia, including the following [52]:

1. An annual mean maximum and minimum temperature dataset consisting of 224 stations, developed by Torok & Nicholls [68]. This dataset was then updated and improved by Della-Marta et al. [69] by re-examining the original set and reducing the number of stations to 133.

2. A daily temperature dataset including 103 stations focusing on the period 1957–1996, developed by Trewin [70].

3. The current ACORN-SAT dataset, consisting of 112 daily records as discussed above, also developed by Trewin [52].

ACORN-SAT has been updated to version 2 at the end of 2018, with reconstruction methods used similar to version 1 [71]. Version 2 incorporates more data, including new measurements and recently-digitised historical records, and also includes further technical updates [72]. Version 2 does however indicate a 23% stronger warming trend than version 1.[12] Comparisons between the raw CDO data and both ACORN-SAT versions were made available for each of the 112 sites at [73].

[12] The mean annual rate of warming for Australia over 1910–2016 is 0.1 °C per decade in version 1, whereas the same rate is 0.123 °C per decade in version 2 as explained on p. 29 of [71].

4 Spatial interpolation techniques

Spatial interpolation is the estimation of attribute values at unsampled locations from measurements made at control points in the same area [74]. The concept is illustrated in Fig. 12, where the value $\hat{z}$ at unsampled location x_0 is to be estimated from neighbouring values $z(x_n)$. Following this approach, several techniques have been developed to reconstruct historical temperature [75], and other environmental variables [76].

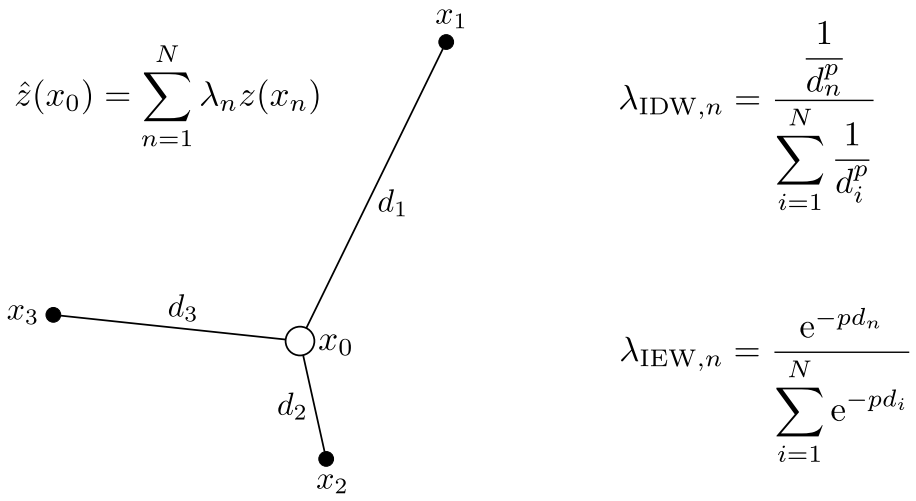


Figure 12: Diagram to illustrate spatial interpolation with distance-weighting formulas.

Spatial interpolation can be performed by calculating the weighted average of neighbouring values, using the general formula [76]

$$\hat{z}(x_0) = \sum_{n=1}^{N} \lambda_n z(x_n) \tag{1}$$

with $\hat{z}$ the estimated value of a variable located at x_0, z the observed or measured value at location x_n with λ_n the associated weight value, and N the number of neighbours from which samples are taken to perform the estimation. Two example equations to calculate the weight values are shown in Fig. 12. These and other methods are subsequently considered.

4.1 Inverse distance methods

The first law of geography or Tobler's law is *"everything is related to everything else, but near things are more related than distant things"* [77], under the assumption of positive spatial autocorrelation. This law forms the basis of a number of spatial interpolation

techniques, especially inverse distance weighting (IDW). The IDW technique performs estimation using a linear combination of neighbouring sample values to the point that needs to be estimated. The weights can be calculated using

$$\lambda_n = \frac{\frac{1}{d_n^p}}{\sum_{i=1}^{N} \frac{1}{d_i^p}} = \frac{d_n^{-p}}{\sum_{i=1}^{N} d_i^{-p}} \tag{2}$$

with d the distance between location x_0 and x_n, and p a power parameter controlling how much emphasis is given to nearby locations. When $p = 0$, all N neighbours considered are weighted equally ($\hat{z}$ would then simply be the mean of all neighbouring values). As p increases, closer locations contribute more to $\hat{z}$. The power parameter is often chosen as $p = 2$, with IDW then referred to as inverse distance squared (IDS). It may also be possible to find an optimal value for p, for example when calculating the theoretical IDW performance limit over a range of p-values. In such cases the technique may be referred to as optimal IDW.

By replacing the inverse distance powers in (2) with inverse exponentials, the alternative formulation

$$\lambda_n = \frac{\mathrm{e}^{-pd_n}}{\sum_{i=1}^{N} \mathrm{e}^{-pd_i}} \tag{3}$$

or inverse exponentially weighting (IEW) is obtained [74].

4.2 Geostatistical techniques

Geostatistics is a branch of applied mathematics that has its origins in agronomy, meteorology and mining [76], though it has been applied to various problems where spatially or temporally correlated data is concerned, including real estate valuation [78], temperature estimation [79], and other earth science fields [80].

Danie Krige, a South African mining engineer, developed statistical techniques to estimate the spatial structure of gold ore reserves in the 1950s [81]. The family of geostatistical techniques based on his work was named Kriging, consisting of generalised least-squares regression methods. Kriging can be used to interpolate spatially-correlated variables, using the following process [82, 83]:

1. The spatial dependence between observed values, at a given time instance, is calcu-

lated using the expression known as the experimental semi-variogram

$$\gamma(h) = \frac{1}{2M} \sum_{i=1}^{M} [z(x_i) - z(x_i + h)]^2 \tag{4}$$

where M pairs of data samples are considered. Each pair consists of measurements $z(x_i)$ and $z(x_i + h)$, taken respectively at locations x_i and $x_i + h$. Although the distance between observation points h may consist of a set of predefined values (e.g. when analysing ore reserves in a controlled study where samples are taken at regularly spaced locations), many applications involve non-uniformly scattered data points. In such cases h may be unique for every data pair considered, and values of h should then be combined through binning.

2. A mathematical model is fitted to the experimental semi-variogram, which can be performed graphically by choosing the model that best fits the $\gamma(h)$ vs. h graph. Commonly-used mathematical models include the spherical, exponential, linear, circular and Gaussian expressions.

3. Attribute values at unsampled locations are estimated using the formula given in (1), where the weights λ_n are calculated by solving a system of linear equations (the Kriging system) containing the mathematical semi-variance model.

Several approaches have been developed based on the above description including ordinary Kriging, simple Kriging, universal Kriging, co-Kriging, Kriging with a trend, block Kriging, factorial Kriging, dual Kriging, etc. [76].

4.3 Graphical techniques

Spatial interpolation can be performed by partitioning the areas around observation locations graphically, and then assigning values at unsampled locations according to the partition. An example of graphic partitioning is Voronoi (also known as Thiessen or Dirichlet) polygons [84], illustrated in Fig. 13, where a region is divided up with one observation point per cell.

All points within each Voronoi cell is closer to the central value (or observation point contained in the cell) than any other observation point. For example, all points within the shaded cell of Fig. 13d are closer to observation point 0 than any other. The simplest graphical interpolation is probably the nearest-neighbour method, where all points within each cell are assigned the central value.

Other graphical or geometric techniques include triangular irregular networks (TINs) and natural neighbours [76].

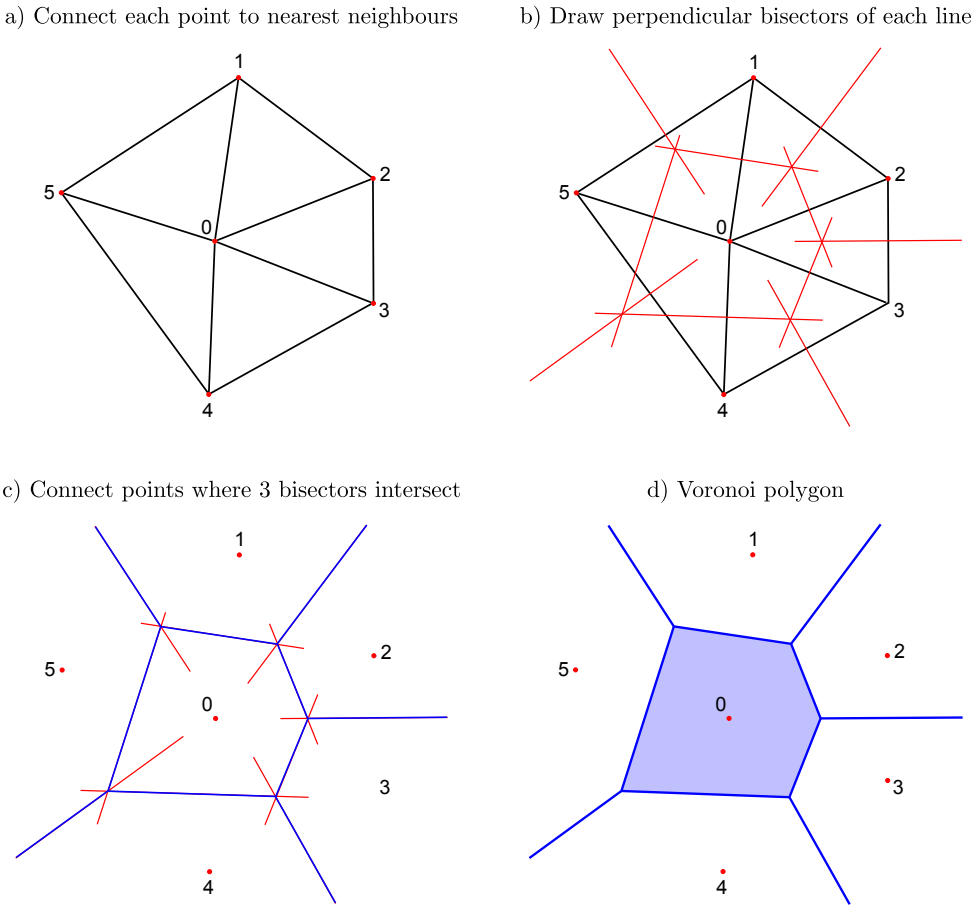


Figure 13: Illustration of drawing a Voronoi polygon around an observational point.

4.4 Polynomial techniques

There are several methods where multivariate mathematical expressions are used to describe spatially-distributed or irregularly-scattered data. Like all spatial interpolation techniques, these methods can either be exact or inexact depending on whether the mathematical expression will reproduce the original sampled data exactly or not. These methods include the following:

- Polynomial regression, e.g. where temperature is assumed to be a third-order polynomial function of the coordinates (x, y) and elevation (z) of the observation site [75]:

$$T(x, y, z) = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 y^2 + b_5 xy + b_6 x^3 + b_7 y^3 + b_8 x^2 y + b_9 xy^2 + b_{10} z \quad (5)$$

Such models are used in trend surface analysis (TSA) where global data trends are described through a regression surface. The parameters of the regression model can be estimated numerically by minimising the residuals using a technique such as least squares.

- Spline interpolation, which can be used to fit curves or surfaces to data points using piece-wise mathematical functions, i.e. a set of polynomial functions is used to describe the area of interest where each function describes a subset of the area with smooth transitions between the different functions [84]. One example of two-dimensional spline interpolation is thin plate splines (TPS), which is a generalisation of one-dimensional cubic splines [85].

4.5 Optimal interpolation

Optimal or statistical interpolation performs estimation of a field using a linear combination of the available data, while minimising the error variance [86]. The formula used is [87]

$$\hat{z}(x_0) = z_f(x_0) + \sum_{n=1}^{N} \lambda_n \left[z_o(x_n) - z_f(x_n) \right] \tag{6}$$

with z_f and z_o the first guess and observed attribute values respectively. Therefore $z_f(x_0)$ is the first guess value at the unobserved location x_0, and $z_o(x_n)$ is the observed value at a neighbouring site located at x_n. The first guess could for example be a corresponding value from the closest neighbour or the neighbour with the highest correlation with the candidate station. The weight coefficients λ_n are calculated by solving a system of equations describing the spatial autocorrelation between all participating stations [87].

4.6 Lapse rate methods

The rate of SAT change in stationary air with increasing elevation is known as the lapse rate, which can be classified as follows [75]:

- Positive lapse rate: Temperature decreases with increasing elevation.
- Zero lapse rate (temperature is isothermal): Temperature remains constant with increasing elevation.
- Negative lapse rate (temperature inversion): Temperature increases with increasing elevation.

Historical temperature data may be estimated for unsampled locations using the lapse rate as follows [75]:

- Calculate the lapse rate using data observed at locations with different elevations in the region of interest.
- Calculate the elevation difference between the unsampled location and the nearest observation site.
- Estimate the unsampled temperature value using the nearest value and the two parameters calculated above.

4.7 Machine learning

Machine learning can be used to perform spatial interpolation by first training artificial neural networks (ANNs) to find a mapping between geographical features and observed attribute values, and then predicting values for unobserved locations using known features of these locations.

Examples of such studies include interpolation of monthly SAT at Mount Kilimanjaro [88] and daily temperature data from the UK [89]. In the UK study an ANN was trained using 34 terrain variables (distances to water bodies, elevation, local roughness, percentage of land cover, urbanisation index, etc.), weather type and temperature observations of 6 neighbouring stations. Temperature surfaces over a 100 km $\times$ 100 km region in Yorkshire were also generated using the trained ANN structure.

An ANN estimation method applied to temperature reconstruction is also presented in Section 8 of this report.

5 Quality control

During the process of measuring and recording weather data, errors may be introduced such that the actual weather is incorrectly represented in the records. Errors may be caused by the following factors [90].

- **Equipment failure**: weather sensors may fail or malfunction due to aging components, extreme weather conditions, electrical or communication faults, etc.
- **Human error**: instruments that are manually operated by human observers may be misread, recorded incorrectly in logbooks or typed incorrectly into electronic databases. Measured values may be recorded at the wrong date or the values themselves may be recorded incorrectly.
- **External influences**: changes in the sensor network and environment discussed in Section 2 could also introduce errors in the data. For example, local phenomena such as bushfire may affect a limited number of weather stations, resulting in an incorrect representation of the regional atmospheric conditions. Other factors, e.g. the UHI effect or vegetation growth near a weather station, could also introduce errors into the data.

Data errors can be identified and removed (or corrected) through a QC process. A number of automatic QC checks can be performed as a first step to identify these errors. QC checks may include the following [90, 91]:

- Climatology checks, which determine whether an observation is within physical limits (e.g. air temperature within -80 and 60 °C) and also test the plausibility of the observation at the given time of the year.
- Internal checks, which compare an observation of a given station with other observations within the same time frame to ensure consistency over time.
- Checking consistency with neighbouring or closest surrounding stations.

The BoM approach to dealing with data values that failed one or more of the above tests, is to place these values on a priority list for a skilled QC operator to investigate further [90]. Information from other sources (e.g. type of measuring equipment used, site location, satellite and radar images, weather charts, etc.) is then used to decide what the relative quality of these potentially wrong observations is. If sufficient evidence exists that

an observation is indeed wrong, it will either be removed or amended, with an appropriate QC flag added.

To address problems in the data caused by changes in the sensor network (e.g. site moves), the BoM applies homogenisation techniques as described in Section 3. Concerns regarding the homogenisation process and how it is being implemented have however been raised. See for example [45], where alternatives are also considered.

5.1 Control charts

An alternative method to performing QC on temperature data using control charts has been presented in [45, 92]. Control charting is a statistical process control (SPC) method commonly used in manufacturing industries, which evaluates whether a process is in a state of statistical control, i.e. behaving in a stable and predictable manner within natural limits of variation. These limits (the upper and lower control limits) describe the usual range of variation due to background noise (also known as common or chance causes) in the process. Exceeding the limits provides evidence that the process is out of control, indicating that an unusual signal (also known as a special or assignable cause) may be present in the process [93].

A common approach to monitor the quality of manufactured products is to take M batches or groups, each containing N samples of the product at a given time instance from the assembly line, and to then subject these samples to statistical testing. To determine whether the process is in control, a quality parameter is calculated from the sample set and displayed on a graph with the parameter mean μ and control limits, typically $\mu \pm 3\sigma$ with σ the standard deviation of the quality parameter. Typically, the values for μ and σ are unknown, and must be estimated when the process is believed to be in control.

The choice of 3σ to establish the control limits originates from the approach in statistics where three standard deviations from the mean is often chosen to define the confidence interval (CI) for a random variable (RV). The 3σ CI is the theoretical interval which will contain 99.73% of the probability mass of a normally-distributed RV, since

$$\int_{\mu-3\sigma}^{\mu+3\sigma} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 0.997300 \tag{7}$$

Although the 99.73% figure only holds under the normal assumption, the Chebyshev inequality states that for a wide class of distributions, at least $100\left(1-\frac{1}{k^2}\right)$ % of observations will fall within the $k\sigma$ CI [94]. Hence, at least 88.89% will fall within 3σ from the mean

and the closer the distribution is to being normal, the closer this CI will approach 99.73% of all observations.

A number of control (or Shewhart) chart types is subsequently discussed. Typically, a combination of control charts will be used concurrently to monitor a number of parameters describing the quality of the outputs of a process, e.g. mean value and variability.

5.1.1 Range control chart

Process variability can be monitored using the control chart for range or the R control chart [93]. In QC applications the standard deviation is often estimated using the range method, with the range of the i^{th} sample group of quality parameters given by

$$R_i = \max(\mathbf{x}_i) - \min(\mathbf{x}_i) \tag{8}$$

which is simply the difference between the largest and smallest samples per group. The average range is then

$$\mu_R = \frac{1}{M}\sum_{i=1}^{M} R_i \tag{9}$$

There exists a well-known relationship between the range and standard deviation of samples from the normal distribution, which has been developed using the relative range defined as [93]

$$W = \frac{R}{\sigma} \tag{10}$$

which can be used to estimate the standard deviation as

$$\hat{\sigma} = \frac{R}{E[W]} = \frac{R}{d_2} \tag{11}$$

with d_2 the expected value of W. Tabulated values for d_2 (and other parameters of W, such as the standard deviation d_3) are available in statistics text books for different values of the sample size N (see for example p. 702 of [93]). For $M > 1$, (11) can be written as

$$\hat{\sigma} = \frac{\mu_R}{d_2} \tag{12}$$

with μ_R given in (9) [93]. To calculate the control limits of the R-chart, the standard deviation of the range σ_R needs to be estimated. Starting again from the relative range defined in (10) and using (12), it is clear that

$$R = W\sigma \tag{13}$$

$$\therefore \sigma_R = \sigma_W \sigma \tag{14}$$

$$\therefore \hat{\sigma}_{\mathrm{R}} = d_3 \frac{\mu_{\mathrm{R}}}{d_2} \tag{15}$$

with the standard deviation of W a known constant; $\sigma_W = d_3$, which can be read from statistics tables as mentioned above. Finally, the R-chart control limits can be developed as

$$\mu_{\mathrm{R}} \pm 3\hat{\sigma}_{\mathrm{R}} = \mu_{\mathrm{R}} \pm 3\frac{d_3}{d_2}\mu_{\mathrm{R}} \tag{16}$$

$$= \mu_{\mathrm{R}}\left[1 \pm 3\frac{d_3}{d_2}\right] \tag{17}$$

and by defining the following constants

$$D_3 = 1 - 3\frac{d_3}{d_2} \tag{18}$$

$$D_4 = 1 + 3\frac{d_3}{d_2} \tag{19}$$

which are also tabulated in statistics text books [93], the upper control limit (UCL) and lower control limit (LCL) can be expressed as

$$\mathrm{UCL}_{\mathrm{R}} = D_4\mu_{\mathrm{R}} \tag{20}$$

$$\mathrm{LCL}_{\mathrm{R}} = D_3\mu_{\mathrm{R}} \tag{21}$$

Fig. 14 shows mean and range control charts using the monthly minimum temperature data of Rutherglen. The sample size is $N = 12$ months for every year (data for years with one or more missing values are excluded). For the range control chart, the constant values are $D_3 = 0.283$ and $D_4 = 1.717$ (read from the table on p. 702 of [93]). The control limits are hence

$$\mathrm{UCL}_{\mathrm{R}} = D_4\mu_R = 1.717 \times 13.595 = 23.343 \tag{22}$$

$$\mathrm{LCL}_{\mathrm{R}} = D_3\mu_R = 0.283 \times 13.595 = 3.847 \tag{23}$$

as indicated in the bottom graph of Fig. 14.

5.1.2 X-bar or mean control chart

Control of the process average or mean quality level is usually performed using the $\bar{x}$ control chart. If M groups of samples are available, each containing N observations of the quality variable, the process average can be estimated as

$$\mu = \frac{1}{M}\sum_{i=1}^{M}\bar{x}_i \tag{24}$$

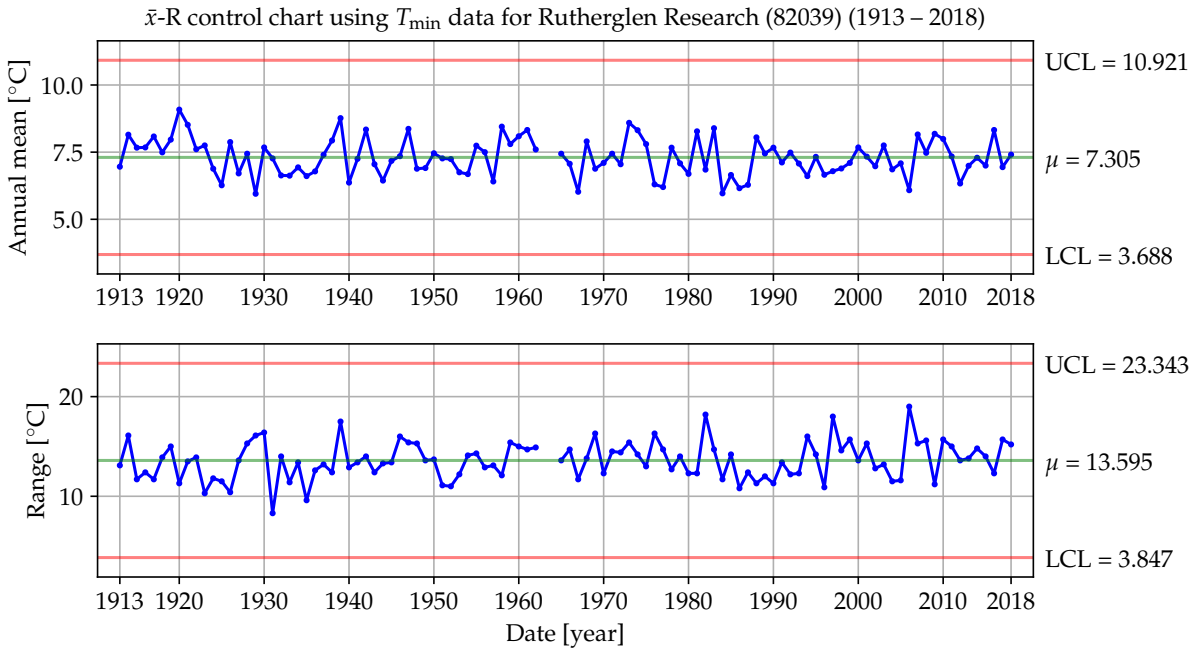


Figure 14: Mean and range control charts for Rutherglen (82039).

with $\bar{x}_i$ the sample mean of the i^{th} group of parameter values. In this case, the sample mean is the variable being subjected to statistical analysis, and the standard deviation of the sample mean σ_μ is therefore required to calculate the control limits. As the mean of each group is calculated over N samples, the standard deviation can be shown to be [95]

$$\sigma_\mu = \frac{\sigma}{\sqrt{N}} \tag{25}$$

with σ the standard deviation of the individual samples within each group (with σ assumed to be identical over all groups as the process is assumed to be in control). The standard deviation σ can be estimated from the average range μ_R as discussed in Section 5.1.1, such that

$$\sigma_\mu = \frac{\sigma}{\sqrt{N}} = \frac{\mu_{\text{R}}}{d_2\sqrt{N}} \tag{26}$$

The control limits can then be calculated as

$$\mu \pm 3\sigma_\mu = \mu \pm 3\frac{\mu_{\text{R}}}{d_2\sqrt{N}} \tag{27}$$

and by defining the constant (which can be read from statistics tables)

$$A_2 = \frac{3}{d_2\sqrt{N}} \tag{28}$$

the control limits can be written as

$$\mu \pm 3\sigma_\mu = \mu \pm A_2\mu_{\text{R}} \tag{29}$$

The top graph in Fig. 14 shows the mean control chart using the monthly minimum temperature data of Rutherglen. The mean value is indicated as $\mu = 7.305$ °C. For the sample size of $N = 12$ observations per group, the constant $A_2 = 0.266$ (read from p. 702 of [93]) and the control limits are therefore

$$\mathrm{UCL}_\mu = \mu + A_2\mu_R = 7.305 + (0.266 \times 13.595) = 10.921 \tag{30}$$

$$\mathrm{LCL}_\mu = \mu - A_2\mu_R = 7.305 - (0.266 \times 13.595) = 3.688 \tag{31}$$

As the control limits are not exceeded, Fig. 14 suggests that the temperature measurement process is in control, and the data are therefore of sufficient quality.

5.1.3 Individuals control chart

If the sample size is $N = 1$ (no grouping), M observed values can be used to construct a Shewhart control chart for individual measurements. The control chart is simply a graph displaying the individual values with the mean and control limits. The top graph in Fig. 15 shows an example individuals control chart displaying the annual average minimum temperature of Rutherglen, suggesting that the process is in control as the limits are not exceeded.[13]

The mean is indicated as the green center line $\mu = 7.305$ °C, which is the mean of all the annual values displayed from 1913 to 2018. To calculate the control limits, the standard deviation is first estimated using (11) with R taken as the average moving range (MR) μ_{MR} between successive annual values, discussed further in Section 5.1.4 below. The control limits can then be calculated as

$$\mu \pm 3\hat{\sigma} = \mu \pm 3\frac{\mu_{\mathrm{MR}}}{d_2} \tag{32}$$

$$= 7.305 \pm \frac{3 \times 0.738}{1.128} \tag{33}$$

$$= 7.305 \pm 1.964 \tag{34}$$

using $d_2 = 1.128$ which is the tabulated value (see for example p. 702 of [93]) when using two samples (or two successive samples in the MR series). The UCL is therefore 9.269 and the LCL is 5.341 as shown in the top graph of Fig. 15.

[13] The data displayed in both top graphs of Figs. 14 and 15 are identical, although the control limits differ. The reason for this is how the data is viewed; for the mean chart in Fig. 14 the average of 12 monthly values is calculated in combination with the range within each group. For the individual chart, each annual value (identical to the mean of 12 monthly values) is viewed as a single value in combination with the moving range between consecutive samples.

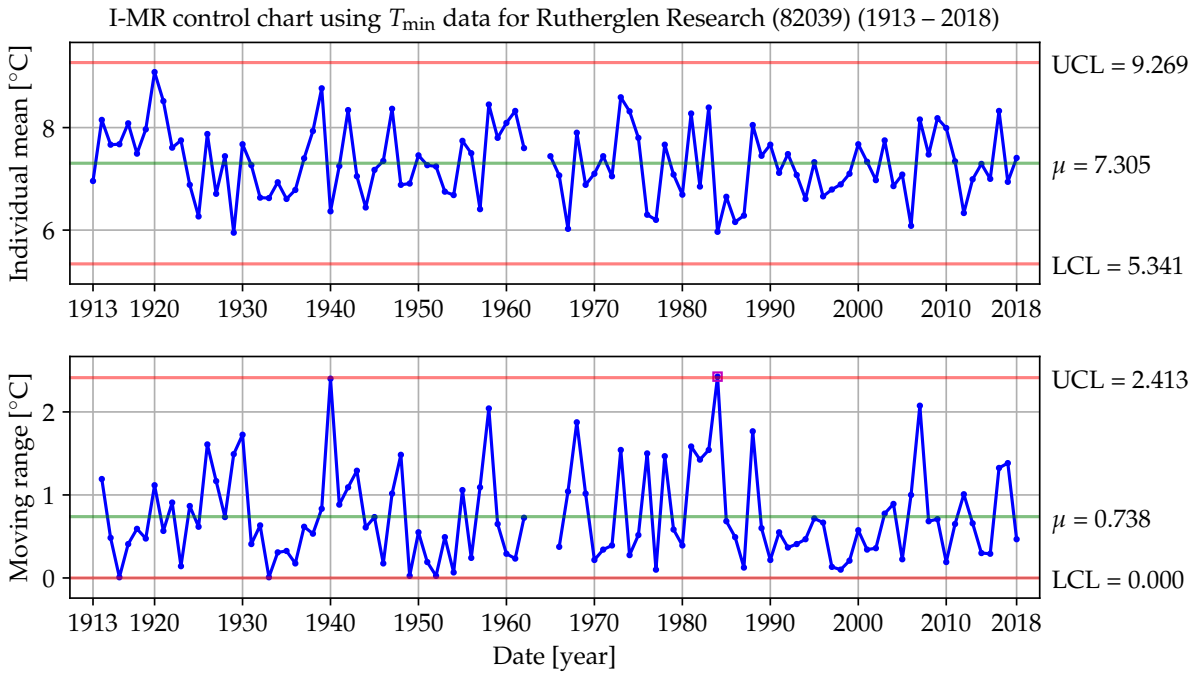


Figure 15: Individual and moving range control charts for Rutherglen (82039).

5.1.4 Moving range control chart

MR control charts are often used in combination with individuals control charts, as the mean MR value can also be used to find the control limits in the individuals chart. Process variability is expressed here in terms of the MR or absolute difference between two successive observations as [93]:

$$\mathrm{MR}_i = |x_i - x_{i-1}| \tag{35}$$

with $i = 2, \ldots, M$. An example MR chart is displayed in the bottom graph of Fig. 15 with the mean MR the average of all MR values calculated as

$$\mu_{\mathrm{MR}} = \frac{1}{M-1} \sum_{i=2}^{M} \mathrm{MR}_i \tag{36}$$

which is $\mu_{\mathrm{MR}} = 0.738$ for the Rutherglen case in Fig. 15. Note that there are two annual values missing (1963 and 1964) and therefore three MR values missing, which have been ignored in the calculations (i.e. no attempt to infill or estimate the missing values is made before calculating the mean or control limits).

The control limits of the MR chart are calculated similar to the R chart, as discussed in Section 5.1.1 with range substituted with moving range. The limits can therefore be

written similarly to (20) and (21) as

$$\mathrm{UCL}_{\mathrm{MR}} = D_4 \mu_{\mathrm{MR}} \tag{37}$$

$$\mathrm{LCL}_{\mathrm{MR}} = D_3 \mu_{\mathrm{MR}} \tag{38}$$

For a sample size of $N = 2$ (or similarly for evaluating MR of span two), the values of the constants are $D_3 = 0$ and $D_4 = 3.267$. The resulting control limits for the bottom graph in Fig. 15 are therefore

$$\mathrm{UCL}_{\mathrm{MR}} = D_4 \mu_{\mathrm{MR}} = 3.267 \times 0.738480 = 2.413 \tag{39}$$

$$\mathrm{LCL}_{\mathrm{MR}} = D_3 \mu_{\mathrm{MR}} = 0 \tag{40}$$

As indicated in the moving range chart of Fig. 15, there are two instances where the control limits are reached or marginally exceeded. There is therefore evidence that the process may be slightly unstable.

5.1.5 Standard deviation control chart

Instead of using the R chart, process variability can be monitored using the control chart for standard deviation or s control chart[14] [93]. The (Bessel-corrected) sample variance of N observations of a RV can be calculated as

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{41}$$

with $\bar{x}$ the sample mean. The sample variance s^2 above is an unbiased estimator of the population variance, that is $E[s^2] = \sigma^2$. However, the sample standard deviation s is not an unbiased estimator of σ. In fact, if the underlying distribution is normal [93]

$$\mu_s = E[s] = c_4 \sigma \tag{42}$$

with c_4 a constant value depending on the sample size, which is available in statistics text books [93]. Furthermore, the standard deviation of s can be expressed as

$$\sigma_s = \sigma \sqrt{1 - c_4^2} \tag{43}$$

[14] The R chart is traditionally more popular due to the ease of calculating the range of a sample set instead of the more computationally-intensive standard deviation. Computer automation has however reduced this effect. Furthermore, the s chart is preferable over the R chart for moderate to large sample sizes [93].

The 3-sigma control limits can then be calculated using (42) and (43) as

$$\mu_s \pm 3\sigma_s = c_4\sigma \pm 3\sigma\sqrt{1-c_4^2} \tag{44}$$

$$= \sigma\left(c_4 \pm 3\sqrt{1-c_4^2}\right) \tag{45}$$

Since σ is typically unknown, it must be estimated from the available data. From (42) it is clear that

$$\sigma = \frac{\mu_s}{c_4} \tag{46}$$

where μ_s can be determined as the average of the standard deviations of M groups using

$$\mu_s = \bar{s} = \frac{1}{M}\sum_{i=1}^{M} s_i \tag{47}$$

with each s_i calculated from (41). The control limits can then further be developed from (45) as

$$\mu_s \pm 3\sigma_s = \mu_s\left(1 \pm \frac{3}{c_4}\sqrt{1-c_4^2}\right) \tag{48}$$

and by defining the following statistics constants

$$B_3 = 1 - \frac{3}{c_4}\sqrt{1-c_4^2} \tag{49}$$

$$B_4 = 1 + \frac{3}{c_4}\sqrt{1-c_4^2} \tag{50}$$

the control limits can finally be expressed as

$$\mathrm{UCL}_s = B_4\mu_\mathrm{s} \tag{51}$$

$$\mathrm{LCL}_s = B_3\mu_\mathrm{s} \tag{52}$$

The standard deviation control chart for Rutherglen $T_{\min}$ is shown in the bottom graph of Fig. 16. The control limits are calculated using (51) and (52) as

$$\mathrm{UCL}_s = B_4\mu_s = 1.646 \times 4.595 = 7.563 \tag{53}$$

$$\mathrm{LCL}_s = B_3\mu_s = 0.354 \times 4.595 = 1.627 \tag{54}$$

Mean control chart using standard deviation

When combining the s chart with the $\bar{x}$ control chart, the control limits of $\bar{x}$ may be calculated as

$$\mu \pm 3\sigma_\mu = \mu \pm 3\frac{\sigma}{\sqrt{N}} \tag{55}$$

with μ and σ the mean and standard deviation of the underlying distribution as before. By substituting σ with the formula in (46), the control limits can then be written as

$$\mu \pm 3\sigma_\mu = \mu \pm \frac{3\mu_s}{c_4\sqrt{N}} \tag{56}$$

and by defining the statistics constant

$$A_3 = \frac{3}{c_4\sqrt{N}} \tag{57}$$

the control limits can finally be expressed as

$$\mathrm{UCL}_\mu = \mu + A_3\mu_s \tag{58}$$

$$\mathrm{LCL}_\mu = \mu - A_3\mu_s \tag{59}$$

which should give the same results as in Section 5.1.2 if the data is normally distributed.

The top graph in Fig. 16 shows the mean control chart for Rutherglen. The control limits for the mean chart are calculated using (58) and (59) as

$$\mathrm{UCL}_\mu = \mu + A_3\mu_s = 7.305 + (0.886 \times 4.595) = 11.376 \tag{60}$$

$$\mathrm{LCL}_\mu = \mu - A_3\mu_s = 7.305 - (0.886 \times 4.595) = 3.234 \tag{61}$$

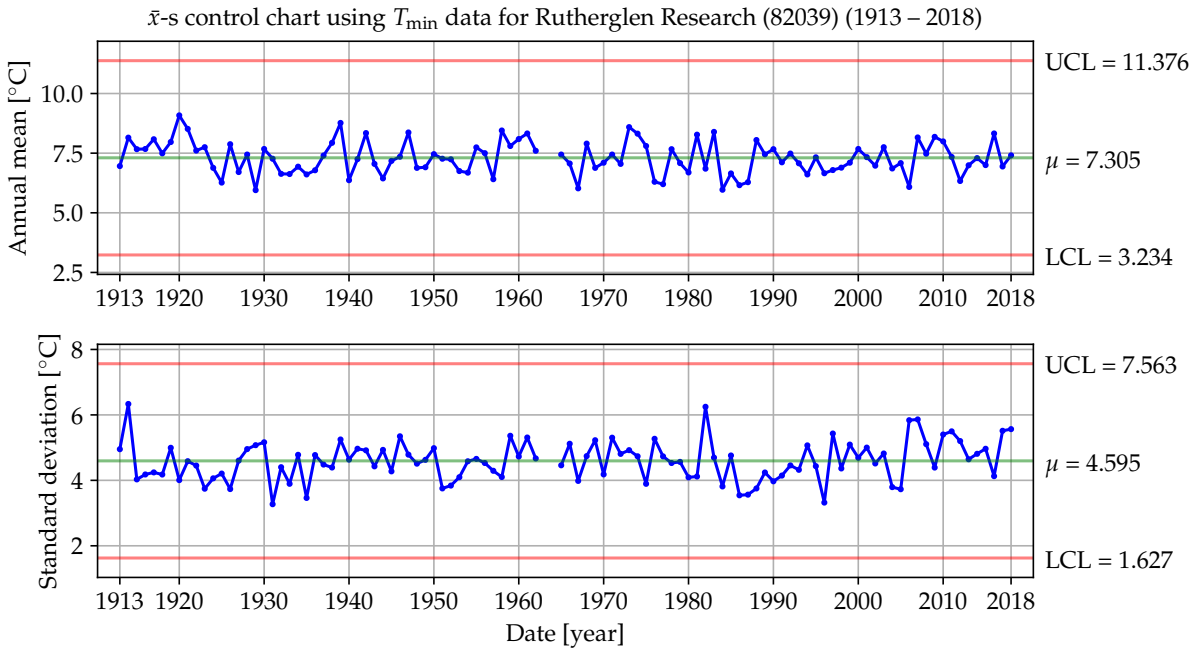


Figure 16: Mean and standard deviation control charts for Rutherglen (82039).

5.1.6 I-MR-S control chart summary

This section provides a summary of the above control chart methods and equations, which can be used to construct I-MR-S control charts to evaluate the quality of monthly temperature data as suggested in [45, 92].

Individuals control chart

1. Calculate the mean temperature value of each complete year from the monthly data to obtain the series
$$\mathbf{a} = [a_1, a_2, \ldots, a_M] \tag{62}$$
consisting of M individual annual mean values.

2. Calculate μ_a, the average of all values in $\mathbf{a}$ formed using (62).

3. Calculate the moving range between successive values of $\mathbf{a}$ using
$$a_{\mathrm{MR},i} = |a_i - a_{i-1}| \quad \text{with} \quad i = 2, 3, \ldots, M \tag{63}$$

4. Calculate μ_{MR}, the average of all values in $\mathbf{a}_{\mathrm{MR}}$ formed using (63).

5. Calculate the control limits of the individuals chart using
$$\mathrm{UCL_I} = \mu_a + \frac{3\mu_{\mathrm{MR}}}{d_2} \tag{64}$$
$$\mathrm{LCL_I} = \mu_a - \frac{3\mu_{\mathrm{MR}}}{d_2} \tag{65}$$
with $d_2 = 1.128$, the tabulated value for using $N = 2$ successive samples in the moving range calculation.

6. Plot $\mathbf{a}$ of (62) with μ_a as the centre line and the control limits given in (64) and (65).

Moving range control chart

As the moving range series $\mathbf{a}_{\mathrm{MR}}$ and its mean μ_{MR} have already been calculated above, the MR control chart can be completed by performing the following steps.

1. Calculate the control limits of the MR-chart using
$$\mathrm{UCL_{MR}} = D_4\mu_{\mathrm{MR}} \tag{66}$$
$$\mathrm{LCL_{MR}} = D_3\mu_{\mathrm{MR}} = 0 \tag{67}$$
with $D_3 = 0$ and $D_4 = 3.267$ for $N = 2$.

2. Plot $\mathbf{a}_{\mathrm{MR}}$ determined using (63) with μ_{MR} as the centre line and the control limits given in (66) and (67).

Standard deviation control chart

1. Calculate the sample standard deviation for each set of $N = 12$ monthly values, using

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{68}$$

with $\bar{x}$ the sample mean of each set, to form the standard deviation series

$$\mathbf{s} = [s_1, s_2, \ldots, s_M] \tag{69}$$

consisting of M individual standard deviation values, one per year.

2. Calculate μ_s, the average of all values in $\mathbf{s}$ formed using (69).

3. Calculate the control limits of the s-chart using

$$\mathrm{UCL}_s = B_4\mu_s \tag{70}$$

$$\mathrm{LCL}_s = B_3\mu_s \tag{71}$$

with the constants $B_3 = 0.354$ and $B_4 = 1.646$ for $N = 12$.

4. Plot $\mathbf{s}$ formed by (69) with μ_s as the centre line and the control limits given in (70) and (71).

Fig. 17 shows the I-MR-S control chart for Rutherglen T_{min}, which is a combination of the graphs shown in Fig. 15 and the bottom graph in Fig. 16. The top graph (individual control chart) tracks the annual temperature change, as the within-group variation is eliminated by averaging monthly data over each year. The middle graph (MR chart) tracks between-year variation, and the bottom graph (standard deviation chart) tracks within-group (within each year) variation [45].

If any of the control limits are exceeded, the entire process can be viewed as out of control. This may imply that the weather measurement process at the given site location does not satisfy the quality requirements and the data record should not be included in further analysis. Alternatively, the control limits can be revised after removing the outlier data. This may however result in new outlying values, and the revision process may need to continue until all remaining values are within limits [93].

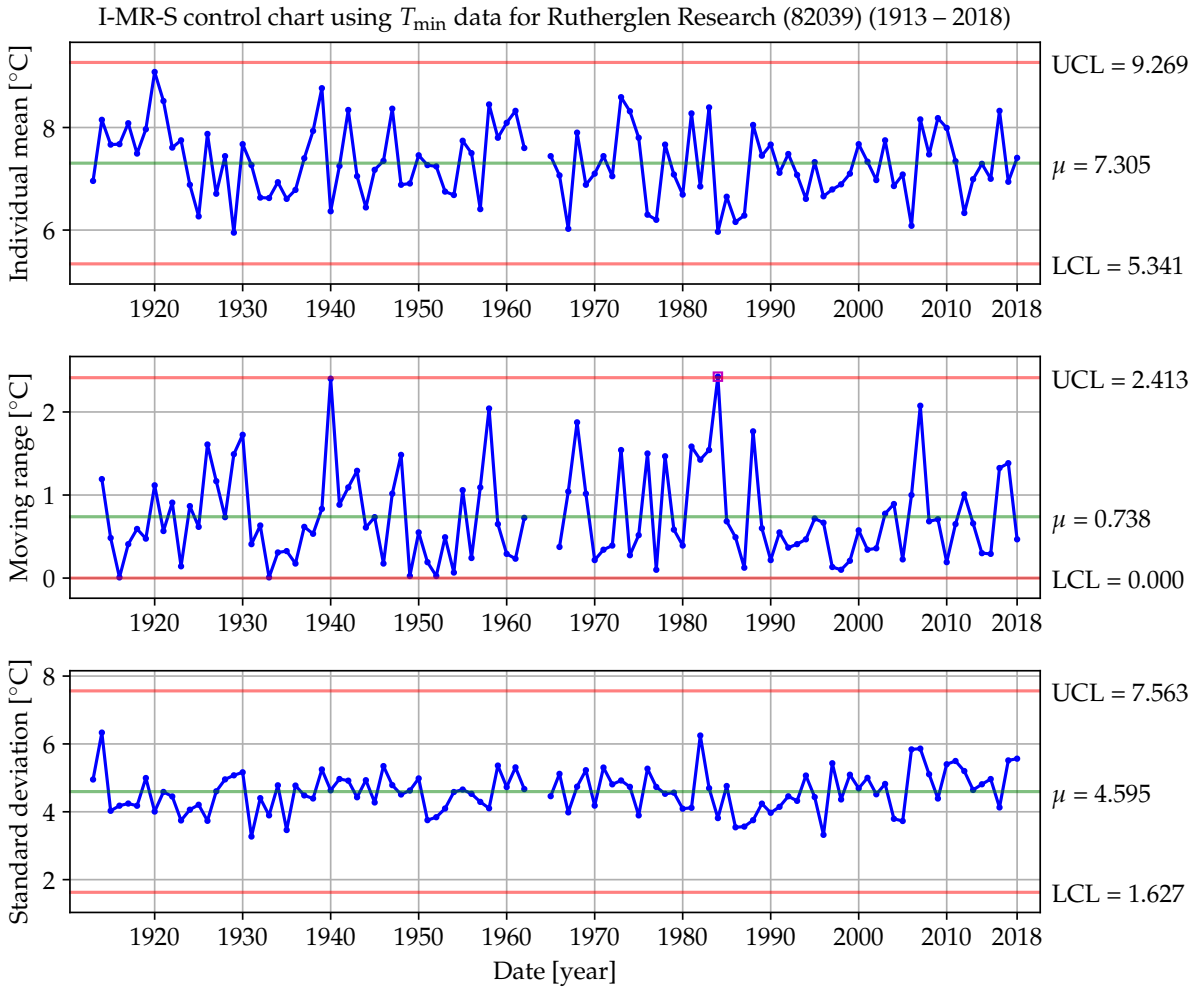


Figure 17: I-MR-S control chart for Rutherglen (82039).

5.1.7 Further examples of I-MR-S control charts

Fig. 18 shows the I-MR-S control chart for Hay T_{min}, an example where several values are out of control. This record seems to be of inferior quality and the record should either be excluded when regional trends are calculated, or an iterative process (excluding single outlier values and revising the control limits) may be used to identify and remove single data points that are problematic.

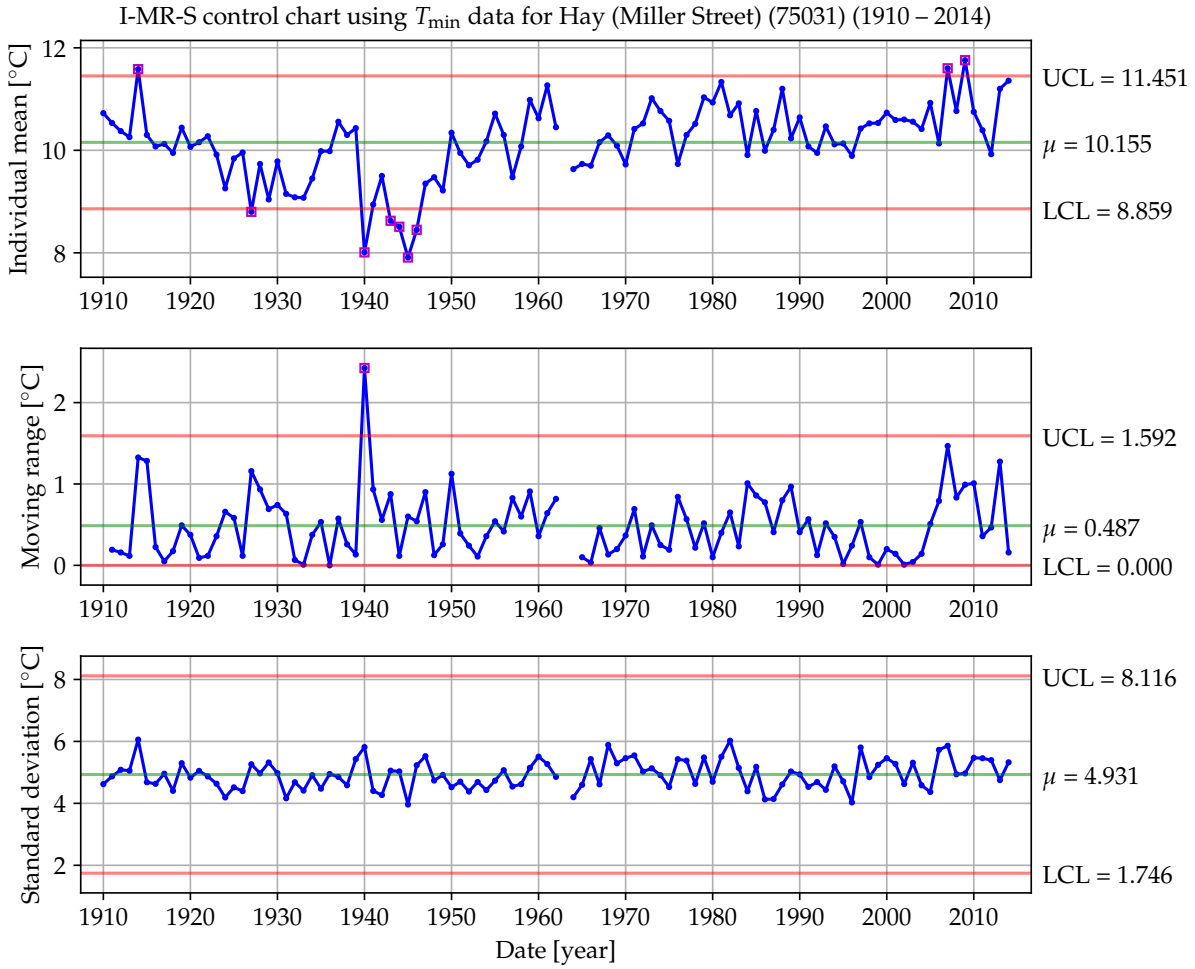


Figure 18: I-MR-S control chart for Hay (75031).

Finally, in preparation for an alternative QC technique presented in the following section, the I-MR-S control chart for Gulpa Island is shown in Fig. 19. The individual chart indicates that 1940 may be a problem, although it seems that the overall process is in control or nearly in control.

As will be shown in Section 5.2, the value for April 1946 is probably spurious, although there is no sign of this in Fig. 19. The possibility of identifying April 1946 as an outlier using the I-MR-S chart is clearly diluted through the averaging process.

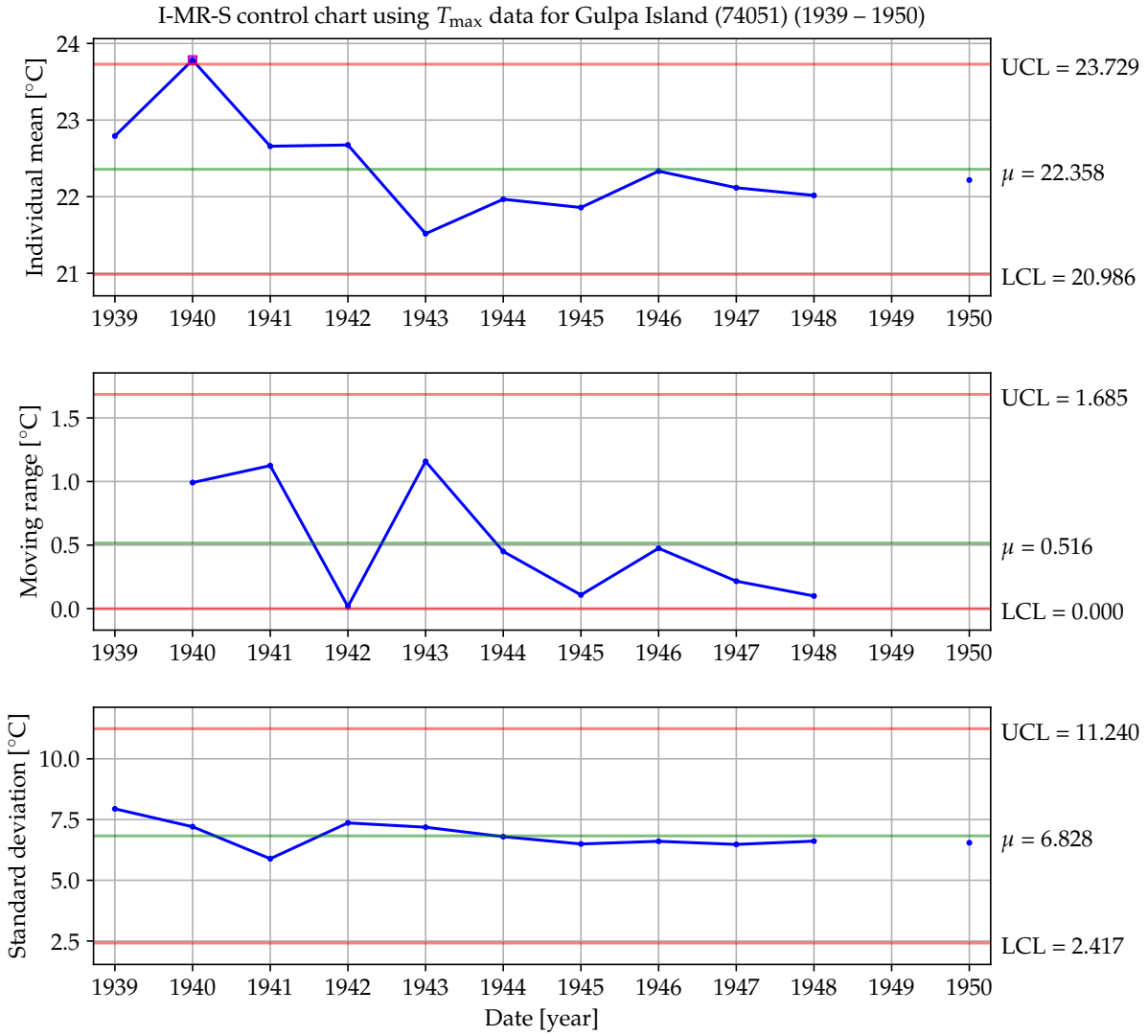


Figure 19: I-MR-S control chart for Gulpa Island (74051).

5.2 Quality control through nearest neighbours

The control chart method discussed in Section 5.1 can be used to evaluate whether single temperature records are of sufficient quality. The method relies on statistical analysis of individual time series to identify outliers, which may be indicators of insufficient quality.

However, outliers may be caused by rare weather events which the control chart (especially individual charts) will view as spurious data. On the other hand, the effect of erroneous spikes may be diluted where averaging is applied, potentially hiding corrupt data. Furthermore, judging the quality of short records is difficult using control charts, and data values not forming part of a full calendar year will be discarded as monthly data are grouped together to calculate annual values.

An alternative QC method relying on data from the nearest neighbouring weather stations is presented here. Instead of using a data series in isolation to evaluate its overall internal quality over time, the quality of single values in the series is determined through comparison with nearby observations made at the same time.

5.2.1 Finding the nearest neighbours

The first step in performing QC using the nearest-neighbour technique is to create a list of neighbours for every station. Every station in Fig. 5 is selected as the central station in turn, and the distance between the central station and all other individual stations is calculated and ranked in ascending order. The distance between any two stations is calculated using the Haversine formula [96], which is the formula for calculating the great-circle distance between two points on a sphere given their latitude and longitude coordinates. As an example, Fig. 20 shows the map of neighbouring stations surrounding Gulpa Island.

5.2.2 Testing potential outliers against neighbours

Fig. 21 shows the monthly $T_{\max}$ series of Gulpa Island, which will be used to illustrate the QC technique. The value for April 1946 is 29.6 °C, which is more than 7 °C above the average for all April months in this record, indicating that this value may be corrupt.

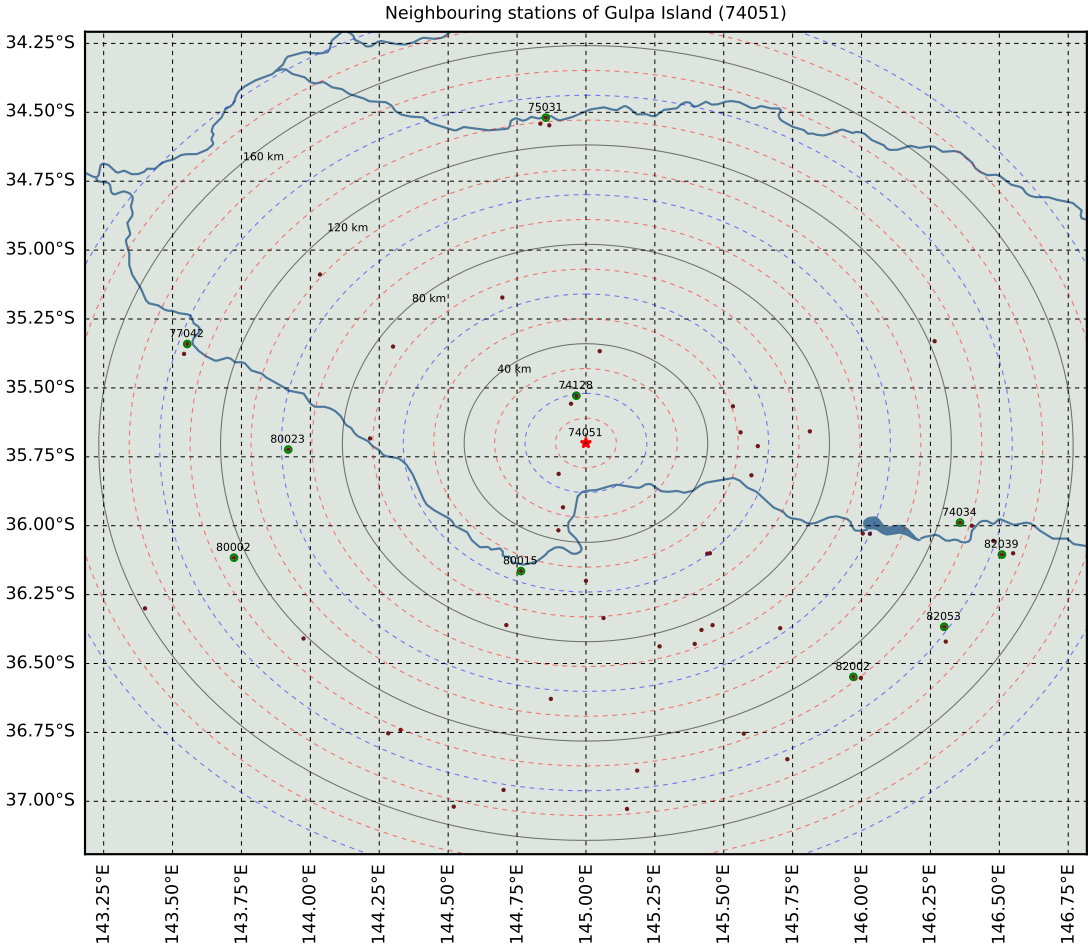


Figure 20: Locations of all SAT-measuring stations active 1910–2018 within 160 km radius from Gulpa Island. The 10 longest $T_{\max}$ records are shown in green with the station ID.

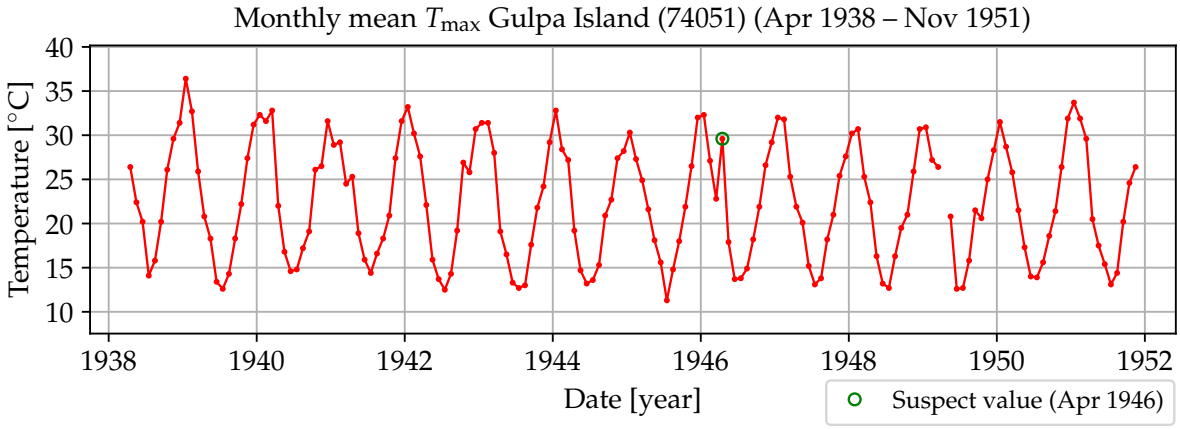


Figure 21: Test series used to evaluate QC method with one suspect value indicated.

To determine whether a single monthly value of the central station record is an outlier, the following process can be followed.

1. Extract the values for the same month from the nearest neighbours of the central station.
2. Arrange all these monthly values in an array, including the monthly value of the central record being tested.
3. Calculate the 3σ CI of the array containing the monthly values, similar to what was done in Section 5.1 using the Bessel-corrected sample variance given in (41).
4. If the monthly value being tested exceeds the CI, it may be flagged as an outlier and excluded in further analysis.

Fig. 22 shows the evaluation of the potential outlier identified in Fig. 21 for April 1946. Twenty neighbours were chosen to calculate the 3σ CI. As shown, the potential outlier exceeds the upper CI limit and is hence confirmed as an outlier that should not be included in further temperature analysis of this region.

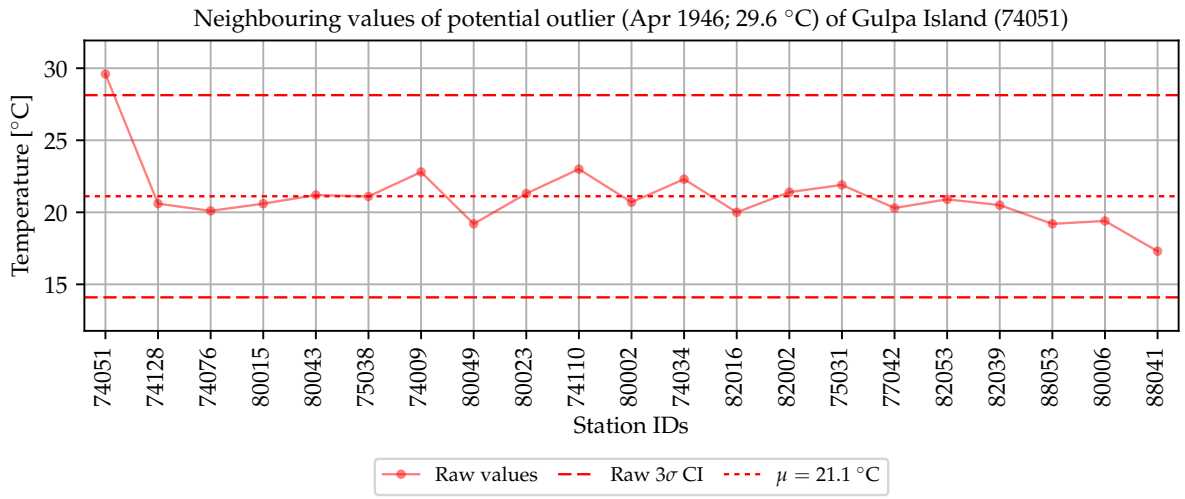
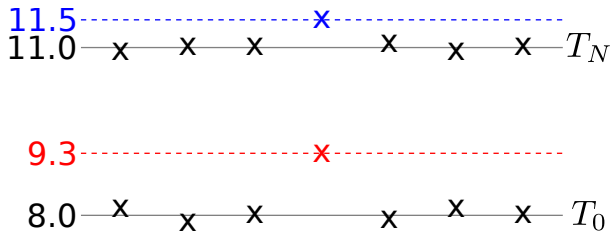


Figure 22: April 1946 $T_{\max}$ value for Gulpa Island and its 20 nearest neighbours.

The above technique of calculating CIs from neighbouring data can be improved by aligning neighbouring data with the central record before calculating the thresholds. It may be possible that a neighbouring location has a consistent offset in temperature, which can easily be corrected using historical data. The correction technique is illustrated in Fig. 23 below, where each neighbouring value being considered is adjusted according to



$$\Delta T = T_N - T_0 = 11.0 - 8.0 = 3.0 \qquad T_{N,\text{rel}} = 11.5 - \Delta T = 11.5 - 3.0 = 8.5$$

Figure 23: Calculation of relative neighbour values.

the difference (or offset) between all other overlapping values between the neighbouring record and the record containing the potential outlier.

Suppose the value 9.3 belonging to the series T_0 has been identified as a potential outlier. Instead of comparing 9.3 with its raw counterpart 11.5 in neighbouring series T_N, the relative offset between T_0 and T_N is first calculated using the 6 overlapping values (black crosses) as $\Delta T = 3$ °C. The relative neighbouring value $T_{N,\text{rel}}$ is then calculated as the difference between the raw neighbouring value and the offset, giving a value of 8.5.

Therefore, with reference to Fig. 23, the simple approach to determine whether the 9.3 value is an outlier is to compare it with raw neighbouring values such as 11.5. The more advanced technique adapts the 11.5 value to 8.5 using the relative offset, such that 9.3 is compared with 8.5 instead of 11.5.

Fig. 24 shows the results when evaluating the outlier of Fig. 21 using the arguably more accurate method. Clearly, the variance of adjusted neighbouring values is less than the raw values, resulting in slightly tighter bounds.

Fig. 25 shows the evaluation of the entire Gulpa Island series shown originally in Fig. 21, when the nearest 20 raw neighbouring values are used to calculate the 3σ CI for each monthly value in the Gulpa Island series.

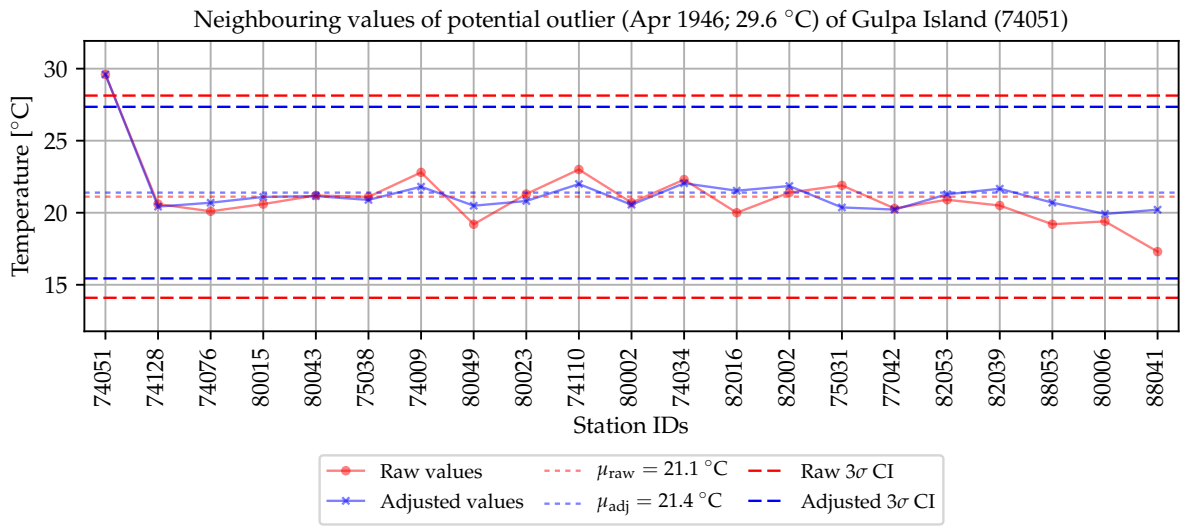


Figure 24: April 1946 T_{max} value for Gulpa Island and its 20 nearest neighbours (including adjusted neighbouring values).

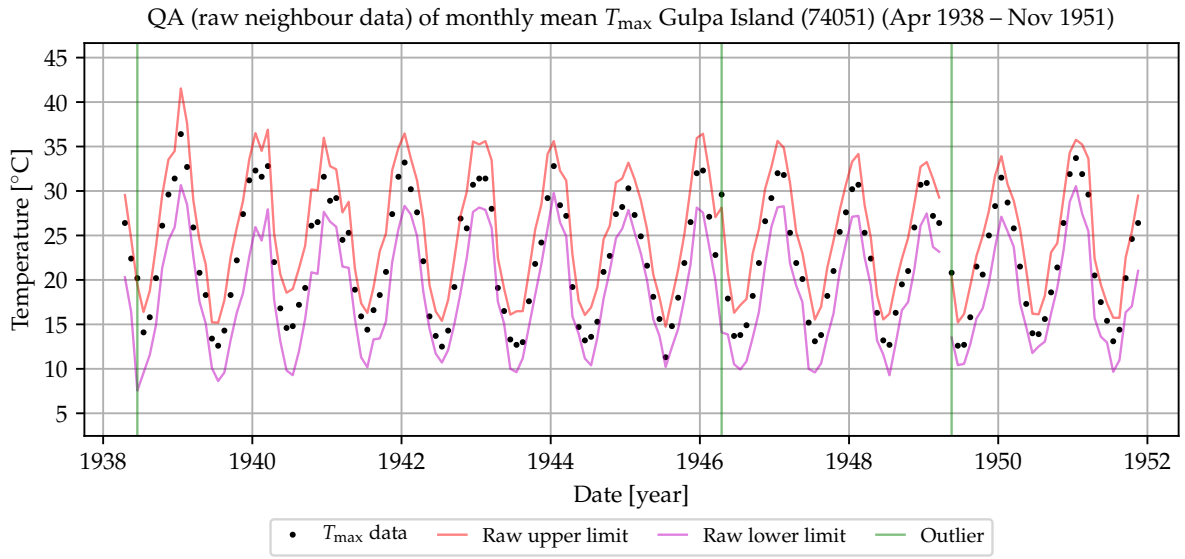


Figure 25: Evaluation of Gulpa Island series using 20 nearest raw neighbour values.

Similarly, Fig. 26 shows the same evaluation when the neighbouring values are first adjusted according to their relative offset with respect to each Gulpa Island value. The raw method resulted in the identification of 3 outliers, whereas the adjusted method resulted in a total of 7 outliers.

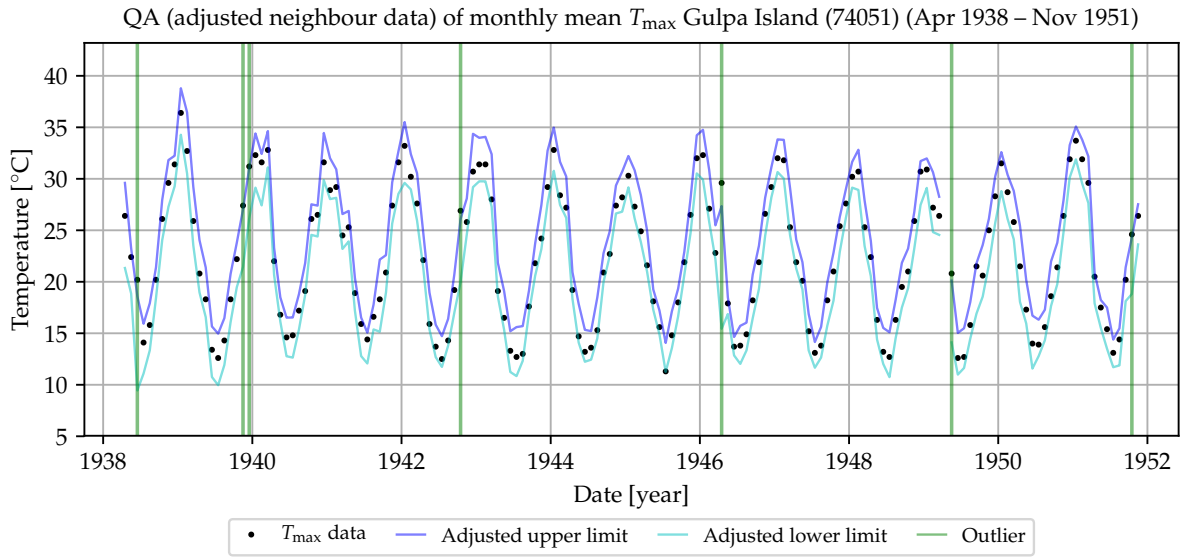


Figure 26: Evaluation of Gulpa Island series using 20 nearest adjusted neighbour values.

5.2.3 Number of samples removed after quality control

There is a total of 725,748 monthly mean T_{max} samples available in the CDO archive over 1910–2018 across Australia. When the two nearest-neighbour QC techniques discussed above are used to evaluate all these samples, and the resultant outliers are removed from every station record, there are 724,843 remaining when using the raw QC technique (see Fig. 25 as example) or 724,472 remaining when using the adjusted QC technique (see Fig. 26 as example).

These and similar values when analysing the monthly mean minimum (T_{min}) record are given in Table 1. The number of neighbours used to identify outliers was set to 20 and the minimum required number of overlapping samples to 6, i.e. there needs to be at least 6 overlapping values in additional to the value being tested, as illustrated in Fig. 23.

	Number of stations	**Number of samples**		
		Raw	**Raw QA**	**Adjusted QA**
T_{max}	1895	725,748	724,843 (905 or 0.125% removed)	724,472 (1,276 or 0.176% removed)
T_{min}	1890	723,288	722,998 (290 or 0.040% removed)	721,344 (1,944 or 0.269% removed)

Table 1: Number of samples before and after performing QC.

Table 1 suggests that only a small fraction of samples is potentially corrupt. The effect of these values will probably therefore not be significant when calculating averages, although they will erode the overall quality of a temperature reconstruction.

6 Nearest neighbour reconstruction

A reconstruction of the historical temperature record of Australia over 1910 to 2018 is presented in this section. All missing monthly values in the raw record are interpolated by infilling the time series of every individual weather station through utilisation of data available at the nearest neighbouring weather stations.

Fig. 27 shows the twelve-month MA of all monthly mean $T_{\max}$ data recorded on the main Australian continent shown in Fig. 5. The number of observations available for every month from 1910 to 2018 is also shown. Of the 1920 stations listed in [36], 1903 are located on the main continent of which 1895 contain data in the 1910–2018 window. A total of 1873 of these records is shown in the top graph of Fig. 27, as they all have at least one continuous 12-month period containing data.

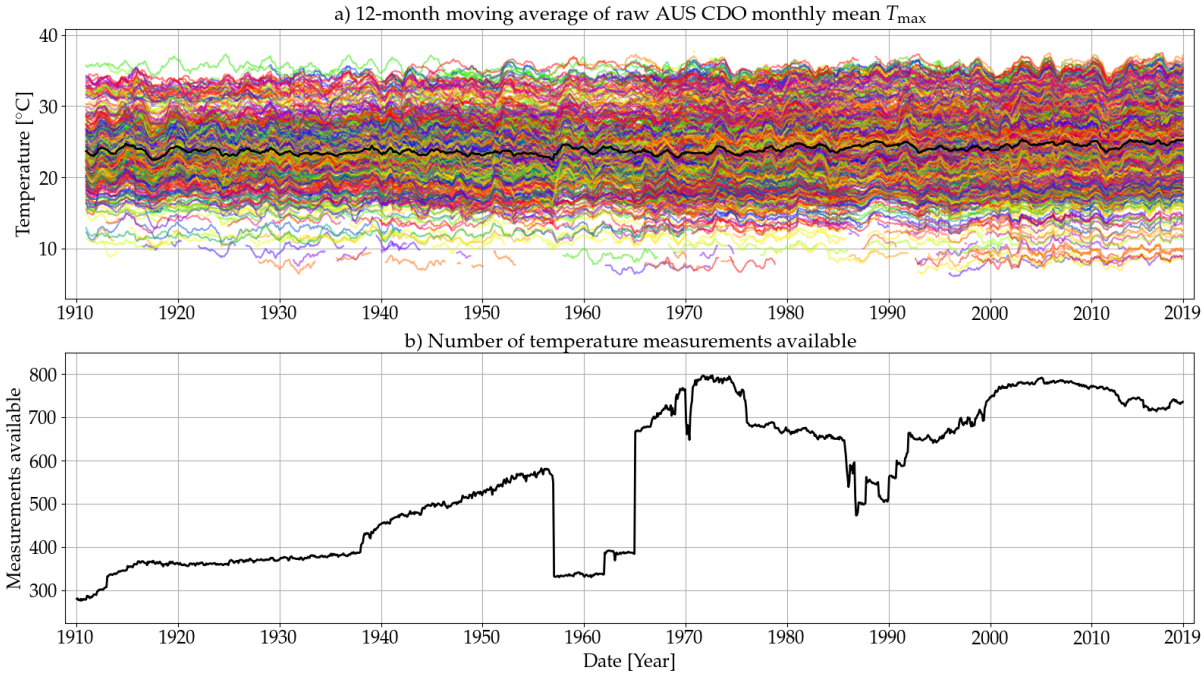


Figure 27: Depiction of all monthly mean maximum temperature data (1910-2018).

The 1895 $T_{\max}$ weather stations on the main Australian continent and nearby islands active over the 1308-month period of 1910–2018, contributed 725,748 monthly observations to the ADAM database. The monthly $T_{\max}$ record can therefore said to be

$$\frac{725748}{1895 \times 1308} = 29.280\% \tag{72}$$

complete over 1910–2018. The aim of the method presented in this section is to increase this figure to 100%.

6.1 Infilling algorithm

The missing monthly data of a given weather station (the *central* station) can be estimated from its nearest neighbours and the relative differences between available temperature data. The algorithm is summarised below, with more details and a step-by-step illustration following thereafter. The algorithm consists of three phases:

1. Preprocessing: Preparing the data and ranking all stations according to the distance between the central station and all other stations shown in Fig. 5, in order to create a list of nearest neighbours.
2. Searching: Identification of the nearest station with sufficient overlapping data with respect to the central station.
3. Infilling: Calculation of the relative offset between data of the central station and the nearest neighbour, and interpolating the missing values.

The preprocessing step is performed only once, whereas the other two steps are performed iteratively for every station, until all missing values are interpolated.

6.1.1 Preprocessing

Preprocessing mainly involves running through all stations in Fig. 5, and constructing the list of neighbours for every station, as was also discussed in Section 5.2.1. Every station is chosen as the central station in turn, and the distance between the central station and all other individual stations is calculated and ranked in ascending order. The distance between any two stations is calculated using the Haversine formula [96], which is the formula for calculating the great-circle distance between two points on a sphere given their latitudes and longitudes.

6.1.2 Searching for the appropriate neighbour

Using the list of nearest neighbours, the station closest to the central station is identified and investigated, which involves determining the number of months of overlap between the two stations. Both stations must at least have a specified minimum number of overlapping months N_{o} containing valid temperature values, before the algorithm will continue to the

infilling stage. If the required number of overlapping months is not met, the next nearest neighbour is identified and the process repeats, until this condition is met.

6.1.3 Infilling

Once a neighbouring station with sufficient overlap is identified, the following process is executed to infill some or all of the missing values of the central station.

1. The average difference between the overlapping values of the two temperature records is calculated.
2. The temperature record of the neighbouring station is adjusted upwards or downwards using the average difference and the two series are then concatenated together, forming an extended (or infilled) record for the central station.
3. Once the data from the neighbouring station has been incorporated into the central record, the record is flagged such that it will be excluded in future iterations.

If the central record still has missing values after the above-mentioned infilling stage, the algorithm will return to the search procedure described in Section 6.1.2, starting again with the neighbour closest to the central station.

6.2 An interpolation example

The algorithm presented in Section 6.1 is illustrated here using the historical Aberfeldy weather station, which was active from 1969 to 1974. This station was simply chosen as the starting point as it is alphabetically the first entry in the list of stations in Australia [36].

Fig. 28 shows the area around Aberfeldy, a small town in Victoria, approximately 125 kilometers east of Melbourne. The locations of the weather stations are also shown, with the Aberfeldy station (ID 85000) approximately in the middle with the ten closest stations encircled in green.

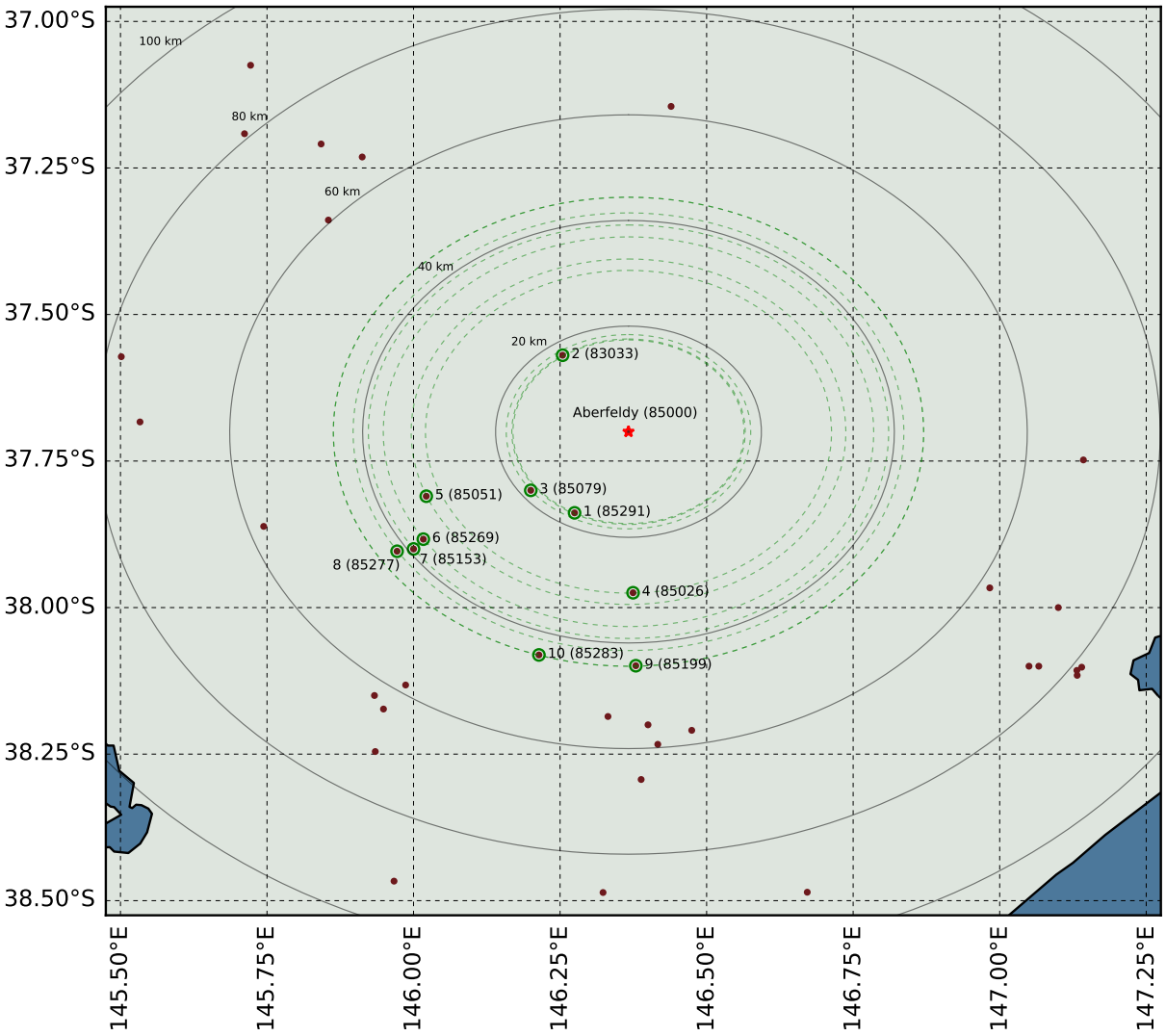


Figure 28: Map of historical weather station locations around Aberfeldy in Victoria, illustrating the ten nearest stations.

6.2.1 Temperature series

The raw CDO monthly mean minimum and maximum series of Aberfeldy are shown in Fig. 29, containing 61 monthly values each - only one value (August 1973) is missing in each series over the available date range. The infilling algorithm must estimate this missing value, and also extend the record backwards to January 1910 and forwards to December 2018.

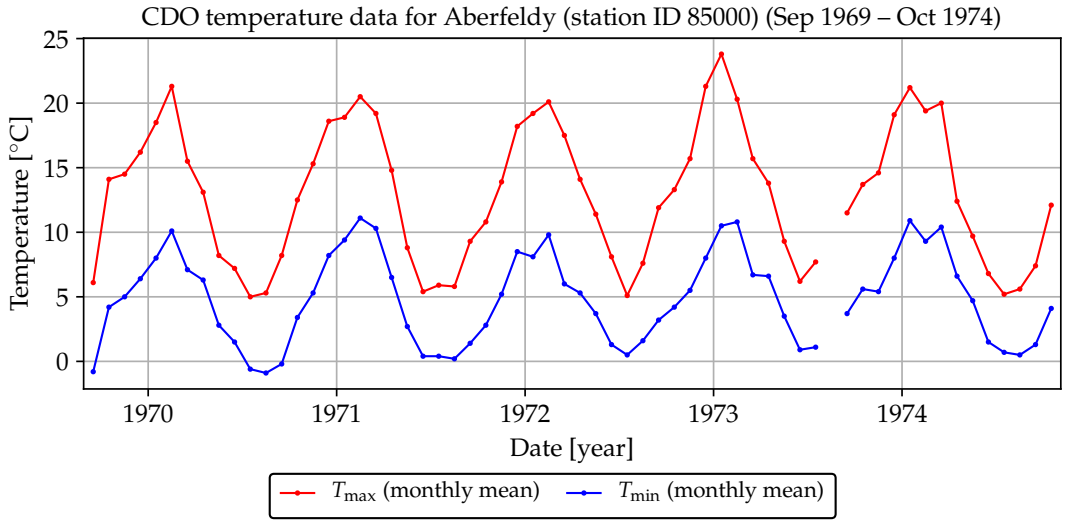


Figure 29: Monthly mean minimum and maximum temperature series of Aberfeldy (ID 85000), each containing 61 values.

6.2.2 Interpolation algorithm

During the preprocessing step, a list of neighbouring stations is created as illustrated in Fig. 28. The searching and infilling steps are then repeated until no missing values are present in the Aberfeldy station record.

Iteration 1

The nearest neighbour of the central station is station ID 85291, as shown in Fig. 28. The monthly mean maximum temperature series T_{max} of these two stations are shown in Fig. 30. By comparing the date ranges of the two records, it is clear that there is no overlap.

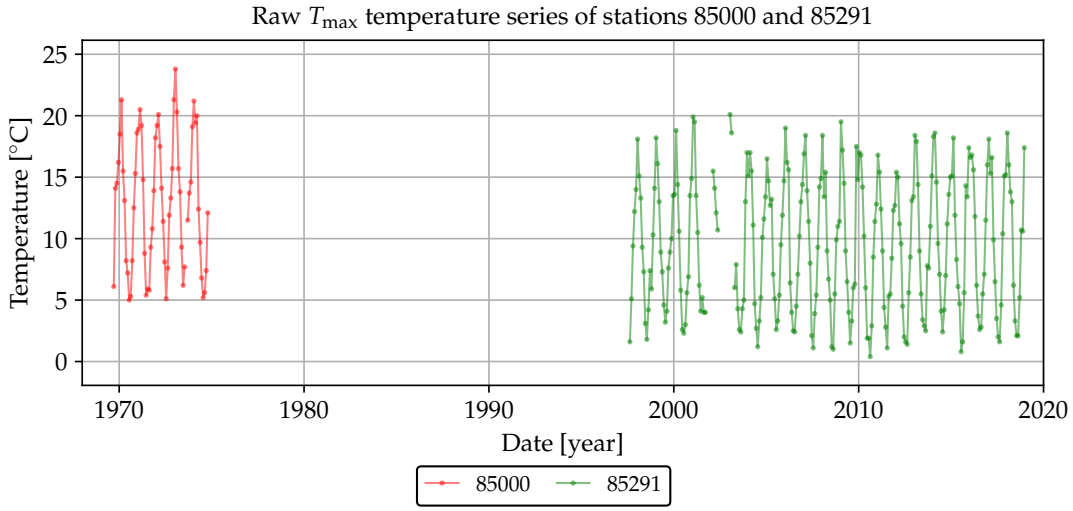


Figure 30: Comparison between the central station and first neighbour, showing no overlap.

The algorithm then moves to the second nearest station (ID 83033), with the $T_{\max}$ series shown in Fig. 31, where there is a one-month overlap in September 1969.

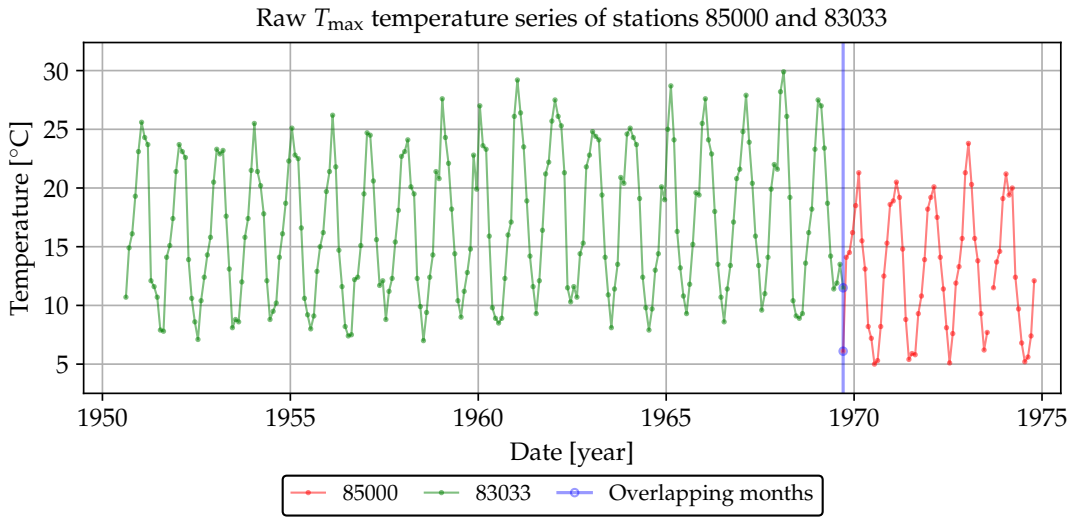


Figure 31: Comparison between the central station (61 values) and the second nearest neighbour (230 values), showing a single-month overlap.

The values for stations 85000 and 83033 for September 1969 are respectively 6.1 °C and 11.5 °C; the average offset is therefore 5.4 °C. If the algorithm accepts a one-month overlap ($N_{\mathrm{o}} = 1$), the data from station 83033 will be imported into 85000. Otherwise, the algorithm will continue searching for a farther station with more overlapping months.

By subtracting the offset value of 5.4 °C from station 83033, the central station can be extended backwards to August 1950, as shown in Fig. 32. Station 85000 has 61 monthly values and station 83033 has 230. The combined record shown in Fig. 32 therefore has $61 + 230 - 1 = 290$ monthly values, as there is only one month overlap.

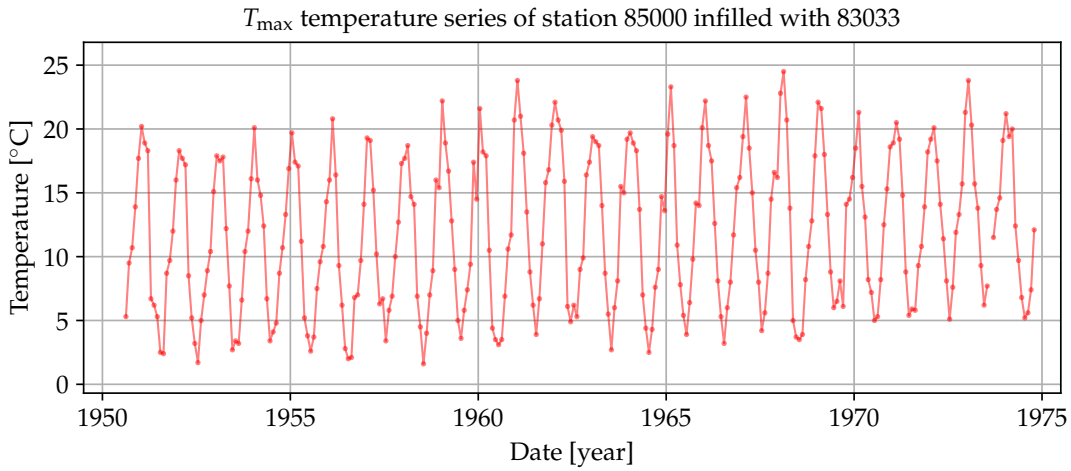


Figure 32: Concatenated record for the central station containing 290 monthly values, consisting of records 85000 and 83033 adjusted 5.4 °C downwards.

Iteration 2

Once the data from the appropriate nearest neighbour is imported into the central record, the algorithm returns to the station closest to the central station, and continues searching for data to import. The concatenated record shown in Fig. 32 is therefore next compared with station ID 85291. The comparison is shown in Fig. 33, with still no overlap, even though the central record has been extended slightly (as a result of the first iteration).

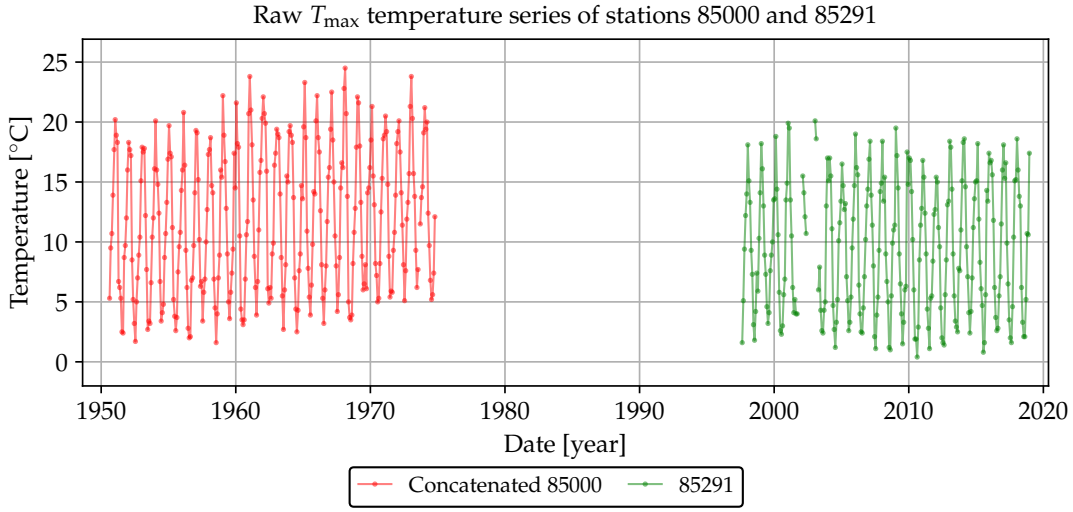


Figure 33: Comparison between the central station (concatenation of 85000 and 83033) and the nearest neighbour (85291), still showing no overlap.

The algorithm continues by comparing the extended central record with neighbouring stations, while ignoring records that were already incorporated. As record 83033 was incorporated into the central record during the first iteration, the algorithm will next compare the extended 85000 record with 85079, the third nearest neighbour. The comparison is shown in Fig. 34, with an overlap of 147 months and average offset of 4.05 °C.

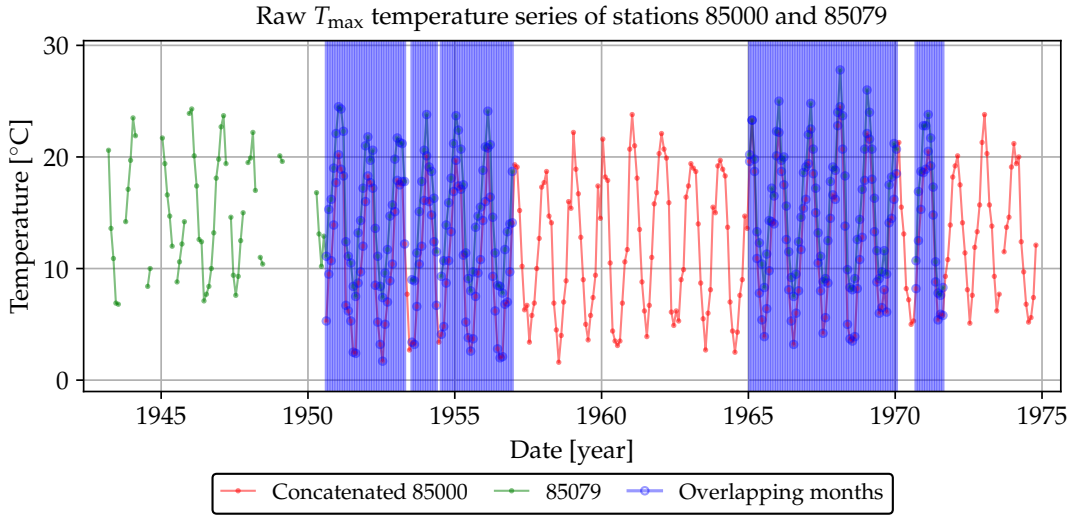


Figure 34: Extended central station and the third neighbour, showing a 147-month overlap.

Station 85079 has 204 values, and there are therefore $204 - 147 = 57$ values that can be added to the central record. By subtracting the offset value of 4.05 °C from station 85079 and extending the central record, the concatenation shown in Fig. 35 is created, consisting of 347 monthly values.

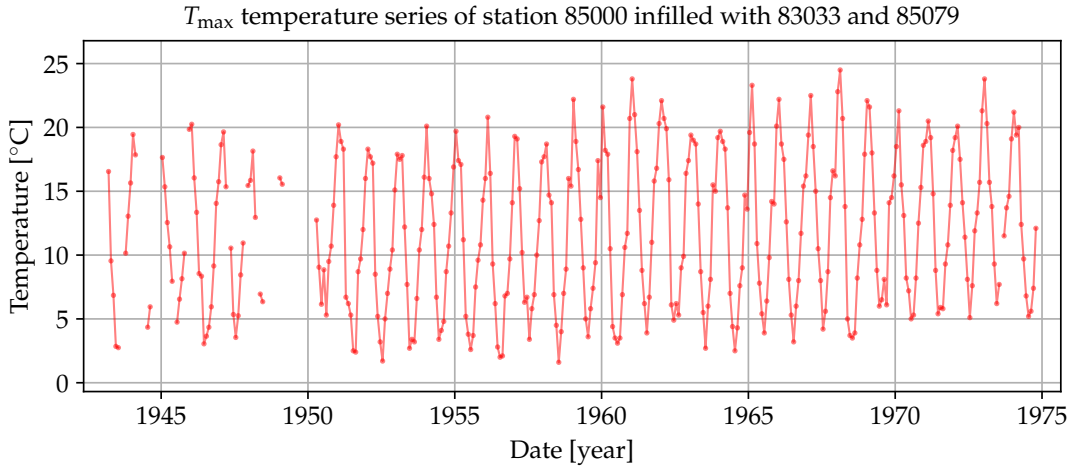


Figure 35: Concatenated record for the central station, consisting of records 85000, 83033 (adjusted 5.4 °C downwards) and 85079 (adjusted 4.05 °C downwards).

This process continues until all missing values are infilled. The result is shown in Fig. 36, which was achieved (for both scenarios shown) by incorporating 18 neighbouring stations into the central record. On average, the T_{max} and T_{min} reconstructions require 13.0 and 14.2 neighbouring stations respectively (for both scenarios shown) to infill all 1895 maximum and 1890 minimum records across Australia[15].

The two scenarios shown in Fig. 36 include choosing the minimum number of overlapping months N_o as 1 and 6 (see Section 6.1.2). The difference between the two is minor in this case, as the minimum number of months is exceeded by a large margin for most station records. Choosing the minimum number of overlap as 6 months instead of 1 will however generally provide improved results, as the offset between two records based on a single month (which does sometimes occur as illustrated in Fig. 31) is not always accurate.

Also note that the minimum number of required overlap should not be chosen too large, as there are a number of short time series in the record. The shortest time series of all stations shown in Fig. 5 is 12 months. By choosing the number of overlap large, less station records will qualify to provide infilling data. For the results shown in Fig. 36 it was however found that choosing the number of overlap larger than 6 has a negligible influence.

[15] Each data record is infilled individually, starting from the scenario where no other record has been infilled.

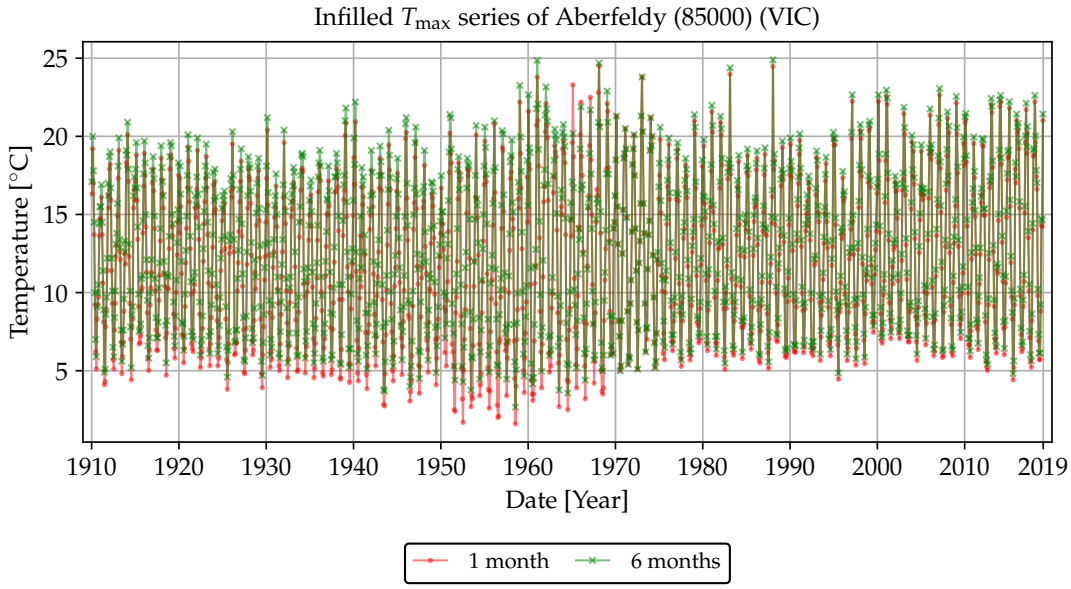


Figure 36: Complete record for Aberfeldy including data contributions from 18 neighbours.

6.3 Nearest neighbour algorithm performance

To evaluate how well missing data is estimated, a performance measure is required. In this section, leave-one-out cross-validation (LOOCV) is used to evaluate the performance of the nearest-neighbour infilling technique, while also evaluating the effect of changes to parameter values (e.g. the number of overlapping months) and the effect of performing QC before infilling the data.

The principle of operation of the LOOCV evaluation technique is to purge a single value from the time series and then to estimate it again. The difference between the true value and the estimated value then provides an indication of the estimation performance of the infilling algorithm. The mean absolute error (MAE) metric is then obtained by running through the entire series and excluding one value during every run, until every single value had a turn to be excluded. A LOOCV-estimated series is then formed by combining all the single estimated values, and the MAE will be the average absolute difference between the original series and the LOOCV-estimated series.

6.3.1 LOOCV results of single series

Fig. 37 shows the LOOCV results for Aberfeldy, including both monthly T_{max} and T_{min}. Every data value was removed from the series in turn and estimated again using the nearest-neighbour infilling technique described in Section 6.1. As shown, the average MAE for the T_{max} series is 0.841 °C and for the T_{min} series 0.513 °C.

The results shown in Fig. 37 are identical for the overlapping number $N_o = 1$ to $N_o = 16$ as the number of overlapping samples between series 85000 and any of its neighbours used here (including 85291, 83033, 85079 and 85026) are never less than 16 for any testing scenario where a single value in series 85000 is excluded. For example, if the minimum overlapping value was chosen as $N_o = 1$, the actual number of overlapping samples will be at least 16 in any case (see Section 6.1.2 for more detail).

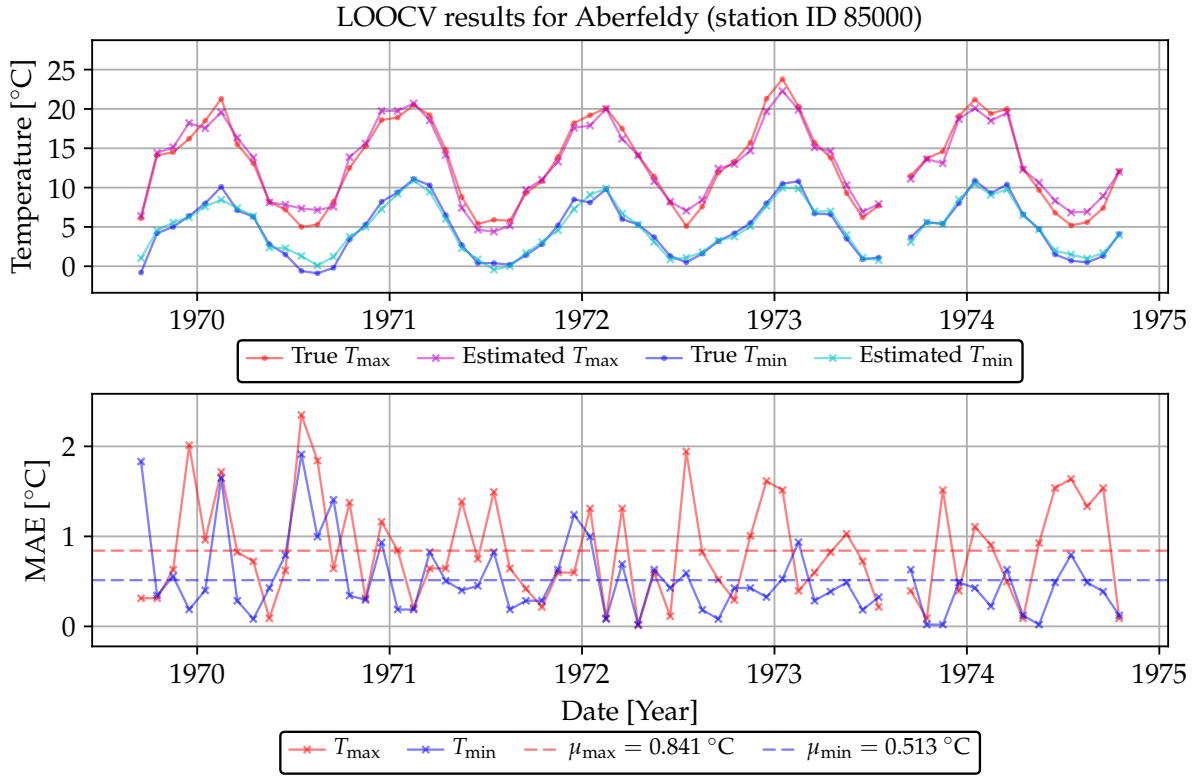


Figure 37: LOOCV results for T_{max} and T_{min} of Aberfeldy (85000).

6.3.2 LOOCV statistics for all CDO temperature data

By calculating the LOOCV-estimated series for every available CDO record and then determining the overall distribution of errors between true and estimated temperature values, the graphs shown in Fig. 38 are obtained.

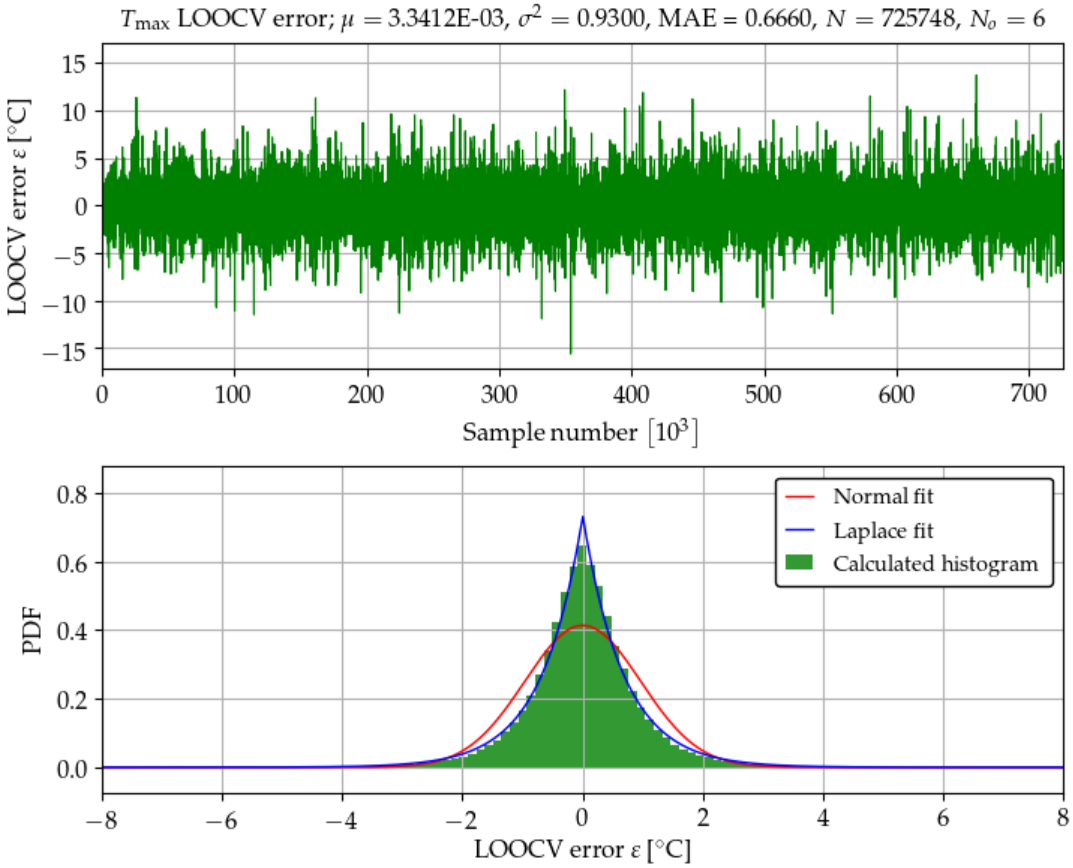


Figure 38: LOOCV time series and distribution for 1895 CDO T_{max} records.

The top graph is the concatenation of the LOOCV error series for 1895 T_{max} records across Australia, containing 725,748 samples over 1910–2018 (see (72)). The bottom graph shows the histogram of this error series and two theoretical probability density function (PDF) curves, including a normal and Laplace distribution fit. Although there are a few very large outliers (errors close to or exceeding 15 °C), the probability mass is still highly concentrated around zero, approximating an exponentially-decaying pattern[16].

The difference in results when using $N_o = 1$ or $N_o = 6$ (the minimum number of required overlapping values to determine the offset between neighbouring data series) is marginal, and both have a MAE of 0.666 °C.

[16] The Laplace density function is $p(x) = \frac{1}{2b} e^{\frac{-|x-\mu|}{b}}$ with $\mathrm{E}[x] = \mu$ and $\mathrm{var}[x] = 2b^2$.

6.3.3 LOOCV statistics for all CDO temperature data after QC

Two QC techniques using neighbouring weather station data were discussed in Section 5.2. These include a raw QC technique where outliers in a temperature series were identified through comparison with e.g. 20 neighbours, and an adjusted QC technique where neighbouring data were first adjusted before comparison with the series being evaluated.

As shown in Table 1 for $T_{\max}$, the raw QC technique identified 905 outliers whereas the adjusted technique identified 1,276 outliers from a total of 725,748 samples. When these reduced datasets are used to perform LOOCV evaluation on all $T_{\max}$ series, the MAE falls from 0.666 (see Fig. 38) to 0.664 (raw QC) or 0.654 (adjusted QC). The results for the latter technique are displayed in Fig. 39.

By comparing Figs. 38 with 39, it is clear that the most extreme outliers were eliminated after QC, as the QC process removed outliers in the original dataset. However, there are still a number of large errors and 10 °C is exceeded twice overall.

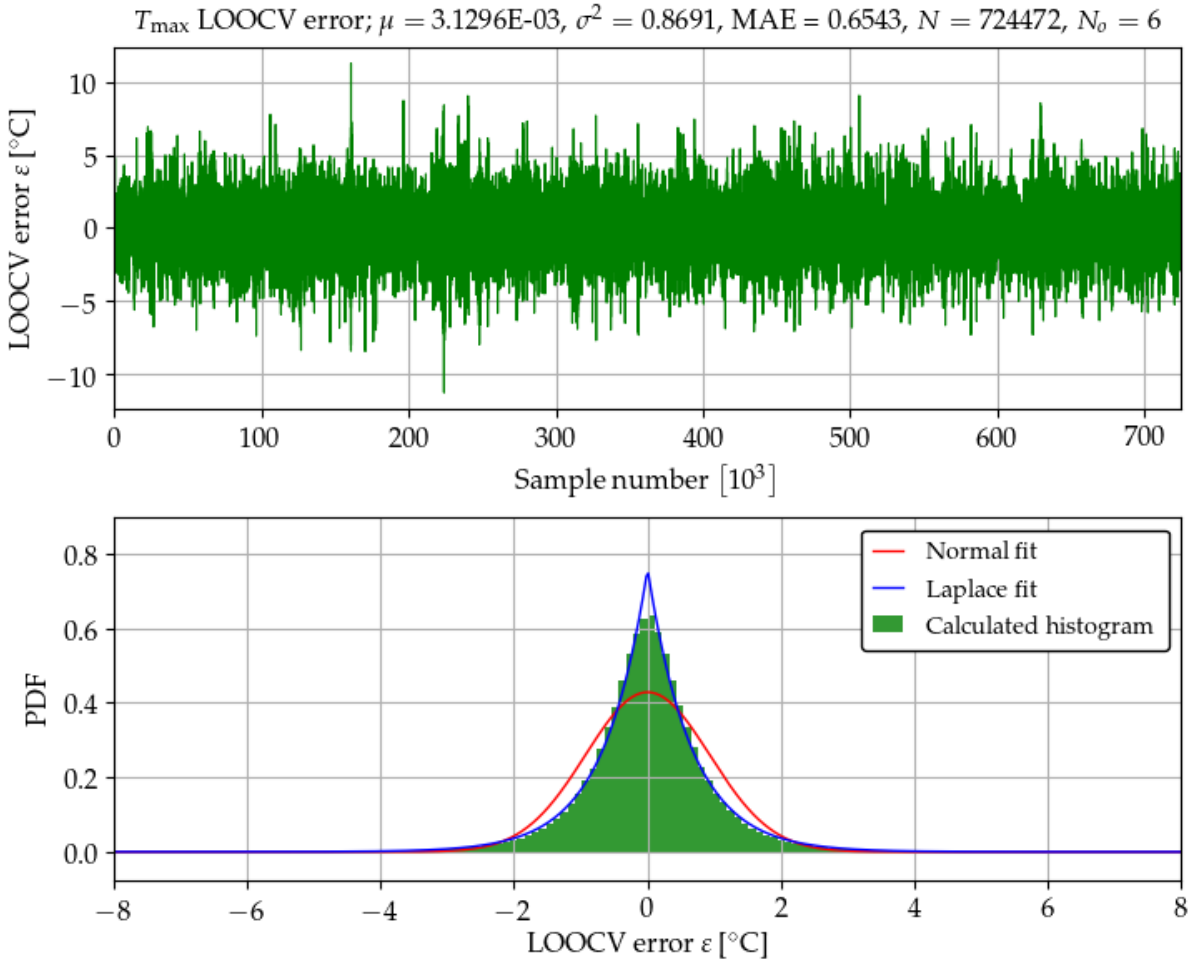


Figure 39: LOOCV time series and distribution for 1895 CDO $T_{\max}$ records after QC.

This largest error exceeding 10 °C is caused by a pair of stations at the same date (Nov 1925), including Cook (ID 18110 at 36.2 °C) and Eucla (ID 11003 at 21.2 °C). The QC process had not identified either value as an outlier, and these two stations are not the closest neighbours (they are separated by 187 km), although for Nov 1925 they are (all nearer stations do not have a value for this date).

This is an illustration of a limitation of this technique; in some cases two neighbouring time series that independently passed QC are not necessarily a good fit to infill missing data between the two, even after adjusting the data according to the average difference between the pair.

Improved infilling performance could therefore be expected by using more neighbouring values (e.g. including more than 1 neighbour in the calculation) and employing additional discrimination techniques (e.g. correlation instead of only separation distance and number of overlapping months) to evaluate whether a given station would be a good choice to estimate missing values of another station.

6.3.4 Infilling the Australian record

If the nearest-neighbour technique described in Section 6.1 is used to infill all the missing monthly $T_{\max}$ values across Australia, Fig. 40 is obtained. The figure shows the twelve-month MA of all individual infilled series, similar to Fig. 27 before infilling.

Each of the 1895 station records were infilled to create an overall record consisting of 2,478,660 samples, which include the original unaltered 725,748 samples. The record spanning 1308 months shown in Fig. 40 is therefore 100% complete. The graphs created when using the raw data after performing QC are similar to Fig. 40 (where the original CDO data were used without performing QC), as the QC process does not affect a large number of samples overall as shown in Table 1.

6.3.5 Comparison with ACORN-SAT

Fig. 41 shows a comparison between the average anomaly series obtained using the nearest-neighbour reconstruction technique (without any QC performed on the raw data) and the ACORN-SAT results originally presented in Fig. 11. The reference period of both anomaly series is 1961–1990.

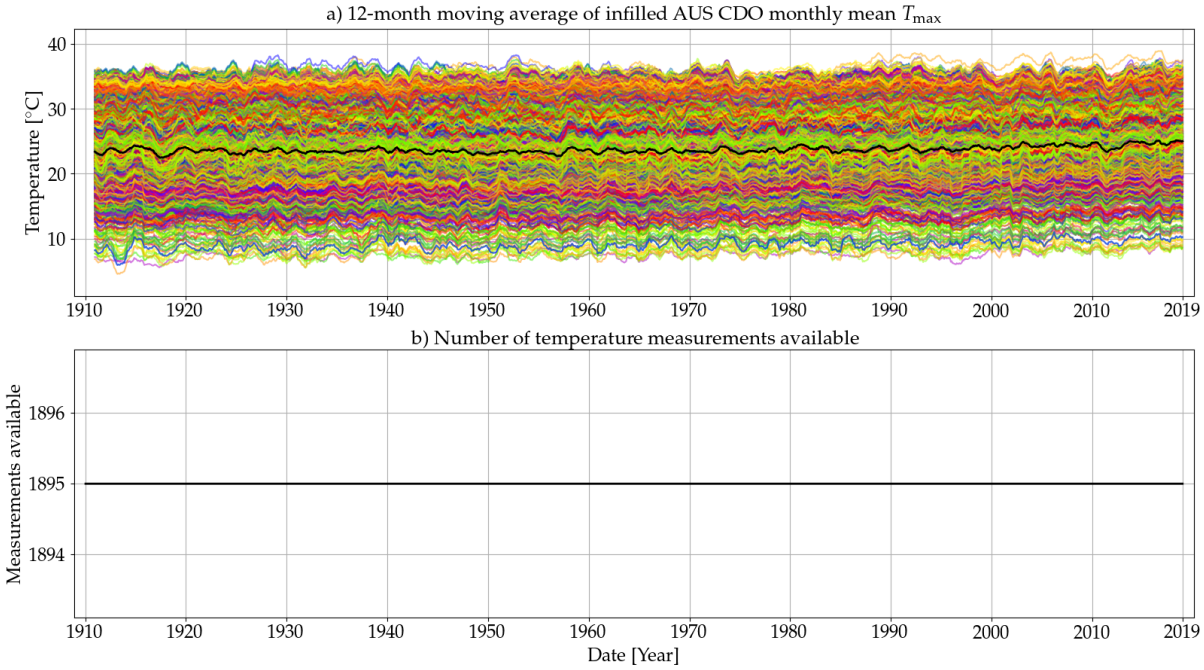


Figure 40: Depiction of all monthly mean maximum temperature data (1910–2018) after nearest-neighbour infilling.

The nearest-neighbour anomaly series shown in Fig. 41 are calculated as follows.

1. All missing monthly data samples of the T_{max} and T_{min} records are infilled using the nearest-neighbour technique described in Section 6.1. The result of this process for the T_{max} record is shown in Fig. 40, which includes 1895 individual station records. Similarly, the T_{min} record includes 1890 individual station records.

2. The annual mean maximum and minimum series are calculated respectively from the infilled monthly mean T_{max} and T_{min} records for each station, simply by averaging the 12 monthly values for every calendar year from 1910 to 2018.

3. The annual mean T_{mean} series for each station is calculated as the average of the annual T_{max} and T_{min} series. There are 1889 individual stations for which a T_{mean} series can be calculated, as there are 6 stations with only a T_{max} (and not a T_{min}) record, and 1 station with only a T_{min} (and not a T_{max}) record.

4. An anomaly series is calculated for every individual annual mean T_{max}, T_{min} and T_{mean} station record where available. Each anomaly series equals the relevant annual mean series with offset removed, where the offset is calculated as the average temperature value over the 30-year reference period 1961–1990. This period was chosen to match the reference period used in ACORN-SAT.

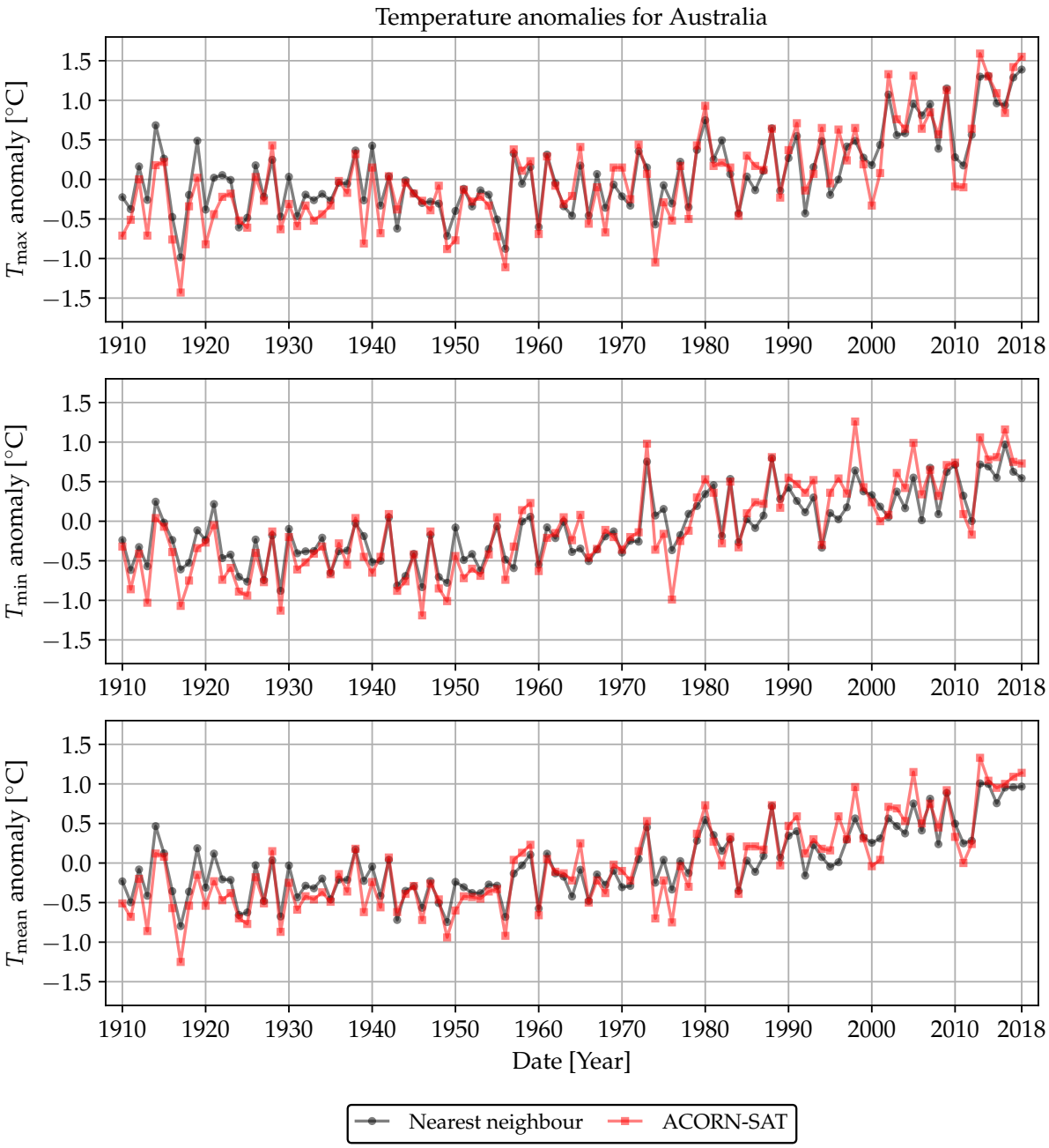


Figure 41: Average anomalies of nearest-neighbour reconstruction vs. ACORN-SAT with reference period 1961–1990.

5. Finally, average anomaly series are calculated to characterise the overall maximum, minimum, and mean temperature patterns for Australia as a whole, which are shown in Fig. 41. Each average anomaly series is calculated as the ensemble average of all available individual anomaly series. That is, an average departure from the reference period 1961–1990 is calculated for every calendar year from 1910–2018 as the average individual departures of single station records for the same year.

Long-term trends and hottest years

Fig. 42 shows the same graphs of Fig. 41 with trend lines and their associated slope values (in °C change over 100 years) added. For the overall period 1910–2018, the ACORN-SAT graphs indicate more severe warming trends compared with the nearest-neighbour reconstruction. For the period 1910–1960, the nearest-neighbour reconstruction shows slight cooling, whereas ACORN-SAT shows slight warming. For the period 1961–2018, both reconstructions show similar warming trends, with ACORN-SAT indicating slightly more severe warming in each case.

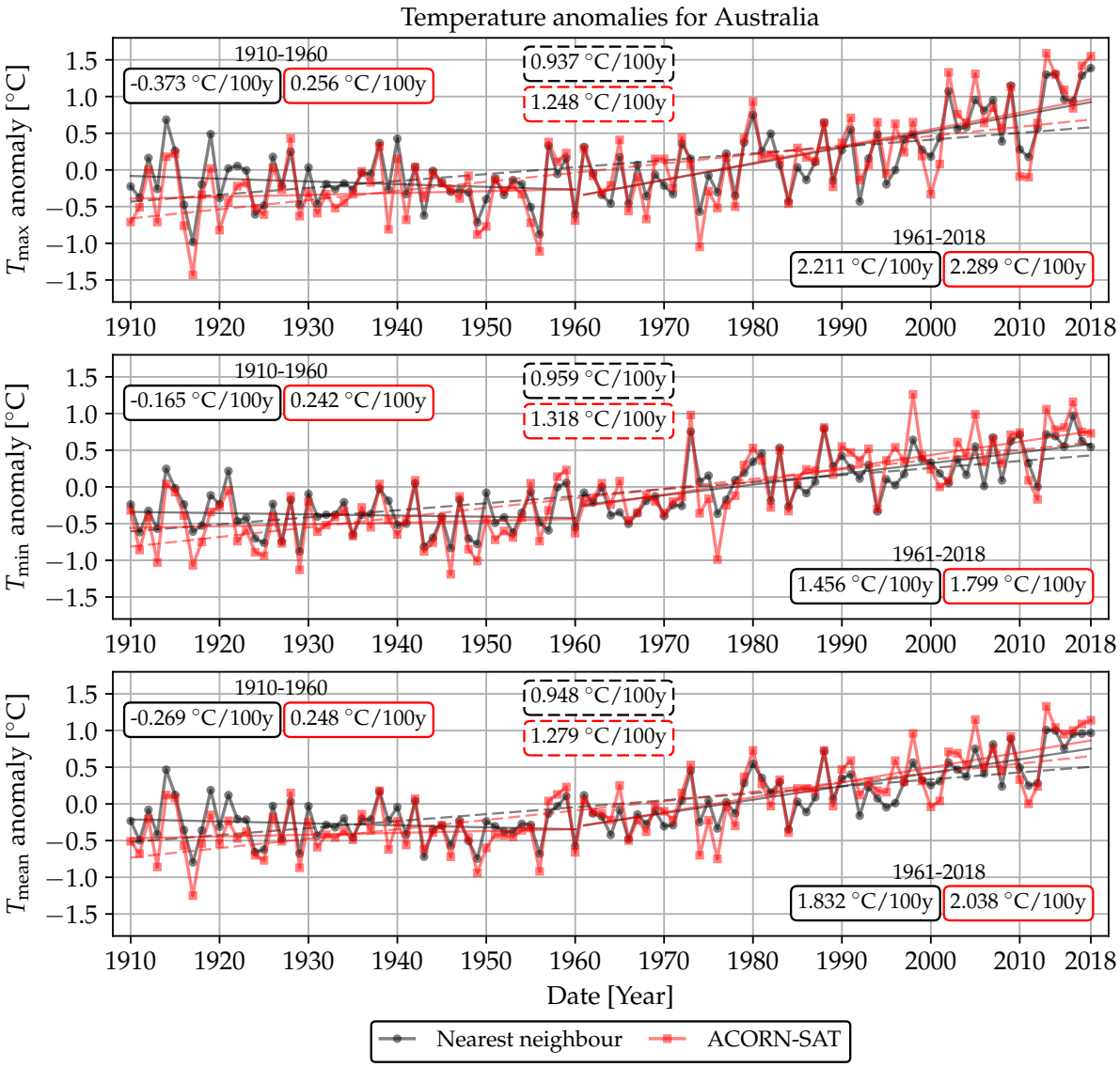


Figure 42: Average anomalies with trends indicated.

The nearest-neighbour results shown in Fig. 42 were obtained using the raw CDO data without performing QC. When the outliers are removed (as discussed in Section 5.2) before infilling all the individual values, the slope values in Fig. 42 are only slightly affected. As mentioned in Section 6.3.4, the QC process does not affect a large number

of samples overall (as shown in Table 1) and therefore the effect on average trends is not significant.[17]

The ten hottest years identified using the average anomaly series shown in Fig. 42 are listed in Table 2 below. Due to the correspondence between the nearest-neighbour and ACORN-SAT $T_{\max}$ reconstructions (especially for the period 1961–2018), both include the years 2013, 2018 and 2017 in the top-4 hottest years (for both $T_{\max}$ and T_{mean}).

	Nearest neighbour						**ACORN-SAT**					
	$T_{\max}$		$T_{\min}$		T_{mean}		$T_{\max}$		$T_{\min}$		T_{mean}	
Rank	**Year**	**Value**	**Year**	**Value**	**Year**	**Value**	**Year**	**Value**	**Year**	**Value**	**Year**	**Value**
1	2018	1.388	2016	0.970	2013	1.007	2013	1.590	1998	1.260	2013	1.330
2	2014	1.309	1988	0.796	2014	1.001	2018	1.550	2016	1.160	2005	1.150
3	2013	1.299	1973	0.753	2018	0.966	2017	1.420	2013	1.060	2018	1.140
4	2017	1.288	2013	0.717	2017	0.956	2002	1.330	2005	0.990	2017	1.090
5	2009	1.150	2010	0.714	2016	0.954	2005	1.310	1973	0.980	2014	1.040
6	2002	1.072	2014	0.691	2009	0.886	2014	1.300	2015	0.810	2016	1.000
7	2015	0.961	2007	0.675	2007	0.813	2009	1.130	1988	0.810	1998	0.960
8	2005	0.955	1998	0.642	2015	0.755	2015	1.090	2014	0.780	2015	0.950
9	2007	0.951	2017	0.625	2005	0.752	1980	0.930	2017	0.750	2009	0.920
10	2016	0.939	2009	0.623	1988	0.720	2007	0.850	2010	0.740	2007	0.750

Table 2: Ranked years according to descending temperature anomaly (10 hottest years).

As shown in Fig. 42, there are a few large discrepancies between the two reconstructions, especially in the $T_{\min}$ graph at 1976 and 1998. The difference between the two reconstructions may be ascribed to the fact that ACORN-SAT is effectively area-weighted, whereas the nearest-neighbour reconstruction only represents the average anomalies of the individual station locations, which are not uniformly distributed as shown in Fig. 5.

This caveat could possibly be addressed by applying spatial interpolation techniques, and extending the infilling technique to unsampled locations (see Sections 4 and 8). To evaluate the effect of using a more uniform station distribution, the nearest-neighbour reconstruction of Victoria is considered next.

Anomaly series for Victoria

The nearest-neighbour anomaly graphs for Victoria are shown in Fig. 43 in comparison with the ACORN-SAT series for Victoria obtained from [67]. These two different reconstruction techniques seem to match fairly well, as both start from the same raw CDO data. The nearest-neighbour technique however does not perform area weighting. In this instance, the need for area weighting is somewhat alleviated by the station distribution.

[17] This however does not negate the need to perform QC, as removing individual corrupt values does improve performance as shown in Section 8.3.

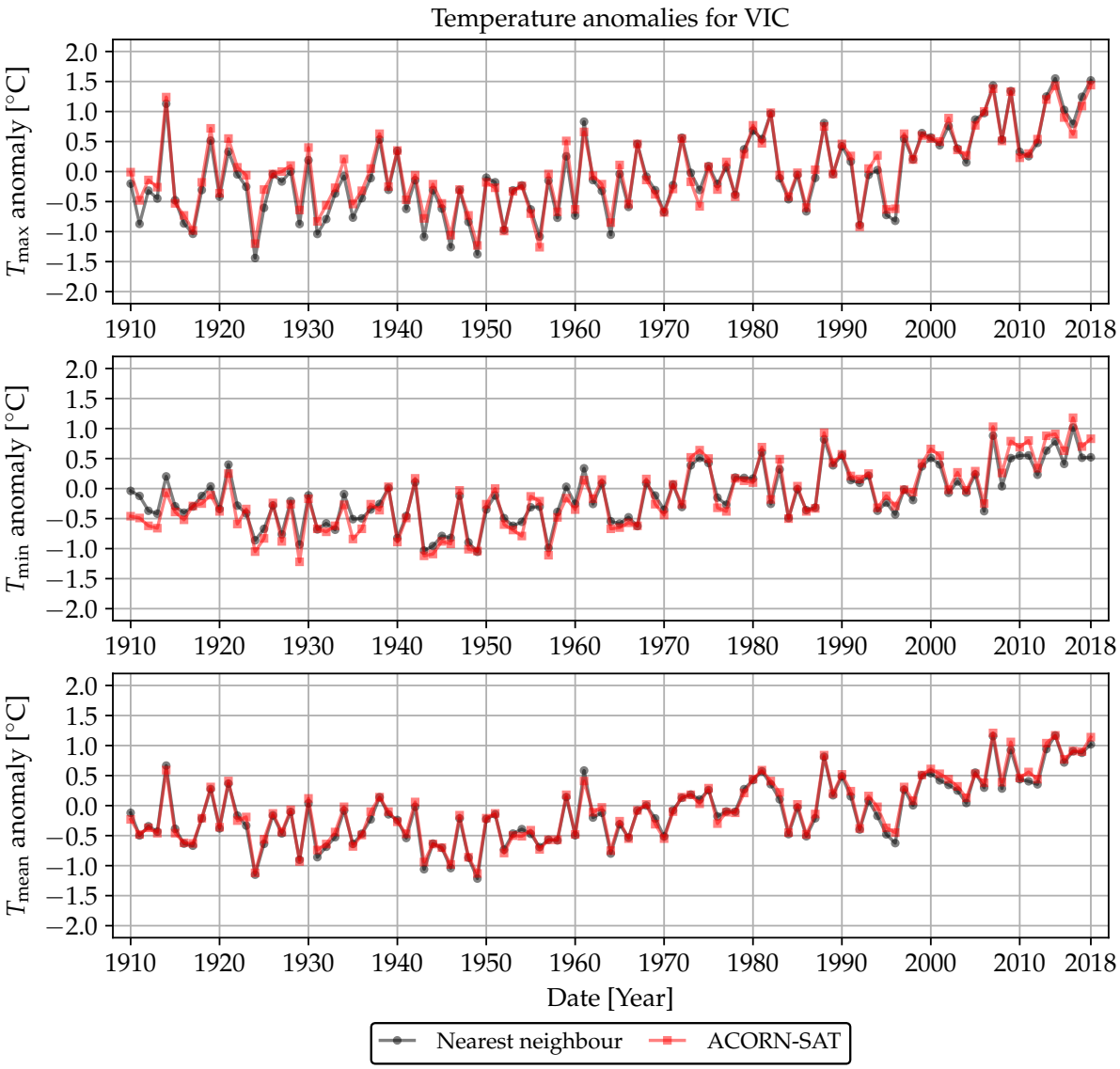


Figure 43: Average anomalies of nearest-neighbour reconstruction vs. ACORN-SAT for Victoria with reference period 1961–1990.

From the Australian map in Fig. 5 it seems that Victoria has the best spatial distribution of stations (of all individual states in Australia) to accurately represent weather patterns across the state. The Victorian stations roughly cover the entire state with only small unsampled locations in between.

Furthermore, by comparing Figs. 41 and 43, it is clear that the anomaly graphs of Victoria alone are closer to the ACORN-SAT result, which suggests the nearest-neighbour reconstruction (for Australia overall and the individual states) could be improved if spatial interpolation techniques are used to include areas without weather stations, when using ACORN-SAT as a reference.

7 Artificial intelligence introduction

Artificial intelligence (AI) can be described as biological or natural cognitive ability endowed to a computing system or machine. Examples of AI applications include optical character recognition (OCR), spam e-mail filtering and autonomous vehicles, which all rely on technologies such as computer vision, pattern recognition and classification.

AI can also be understood in relation with classical computation and the type of problems addressed. Classical computation typically involves the execution of a predefined algorithm or a set of instructions to solve a problem, e.g. finding $\sqrt{2}$ using Newton's method [97]. Such methods are typically intellectually difficult (or tedious) for humans to perform, although not difficult to describe. Computers can easily implement these methods by merely executing a list of formal mathematical rules at high speed.

In contrast, the true challenge of AI is solving problems that are easily performed by humans, although not so easily described in a formal way [98]. For example, humans can intuitively recognise faces, but producing a list of instructions for a computer to perform this task with acceptable accuracy is not easily done.

An often better approach to address such challenges is to allow a computer or AI agent to acquire its own understanding of the problem by exposing it to data. Such an AI agent would then analyse the data, extract patterns, and set up its own implicit rules to solve the problem. This machine learning (ML) capability has enabled computing systems to solve real-world problems with human-like accuracy, and has also led to the adoption of AI in various applications [98].

ML can broadly be achieved through either supervised or unsupervised learning [98]. In supervised learning, the ML system is exposed to several examples of a vector $\mathbf{x}$ (the input data) and an associated value or vector $\mathbf{y}$ (the output, label or target data). The ML system must then discover the mapping between $\mathbf{x}$ and $\mathbf{y}$, in order to predict the output if new input data is presented.

In unsupervised learning, the ML algorithm is not guided through the provision of labelled data, and the algorithm must make sense of the data on its own such as finding hidden patterns or structure in the data. Only supervised learning will however be considered in this report.

7.1 Artificial neural networks

One way of realising AI with ML capability is through a network of artificial neurons, which constitutes a computing framework with the fundamental processing unit a very simple mathematical model of a biological neuron, illustrated in Fig. 44.

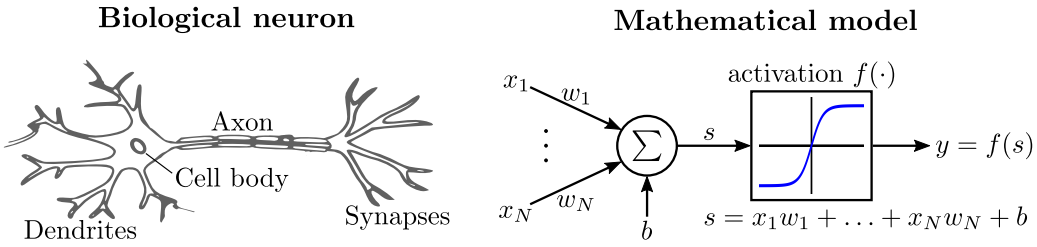


Figure 44: Biological and artificial neuron, which is the basic element of ANNs.

In a biological nervous system, neurons receive, process and transmit information through electrical and chemical signals. Input signals are provided to the neuron via its dendrites, which are then combined and processed via the cell body. If a certain threshold level is exceeded, the neuron "fires" by transmitting constant-magnitude action potentials at a certain frequency through the axon, which then activates synaptic connections with following neurons [99].

As illustrated in Fig. 44, the artificial neuron mimics its biological counterpart by applying a weighting w_i to each input x_i (to assign relative importance levels to each input), summing the result and adding a fixed offset or bias level b, and then transforming this result s using a nonlinear activation[18] or transfer function $f(\cdot)$ to obtain the output $y = f(s)$.

An ANN can be formed by connecting a number of neurons together as shown in Fig. 45. This ANN has three layers, including the input layer, one hidden layer and the output layer. In this example, the input layer contains $N_\mathrm{i} = 3$ nodes (one for each input), the single hidden layer contains $N_\mathrm{h} = 3$ nodes (or neurons), and the output layer contains $N_\mathrm{o} = 1$ neuron. The structure is therefore a N_i-N_h-N_o or 3-3-1 ANN.

The weights w and biases b (collectively *weights*) shown in Fig. 45 are variables, which

[18] The term "activation function" was inherited from the biological neuron's all-or-nothing behaviour; the neuron is either on or off - it activates if the magnitude of the input signals are sufficiently large. Although the artificial neuron does not behave in this way, it could be argued (depending on the ANN activation function) that the artificial neuron is "off" (negative saturation) or "on" (positive saturation) with a graded output between these states.

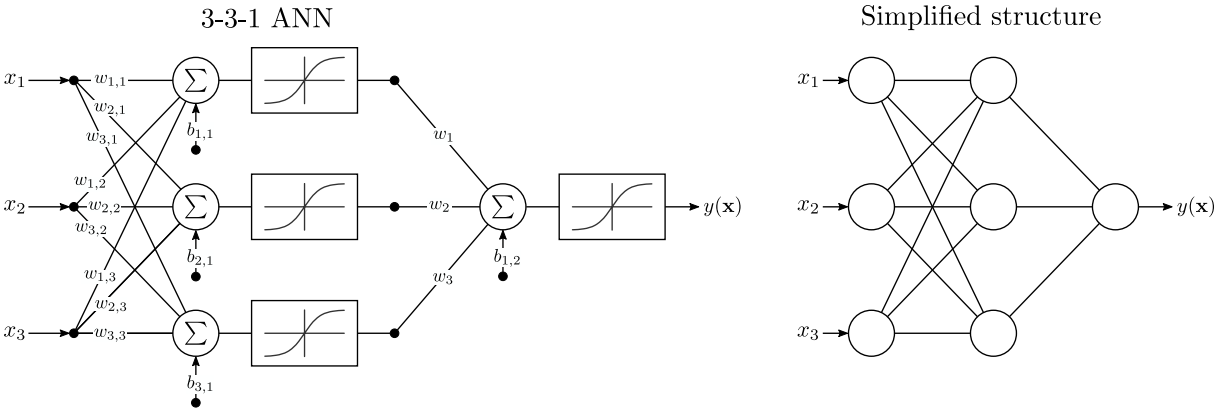


Figure 45: A 3-3-1 ANN structure and its simplified representation.

together constitute the memory of the network. A three-layer network has

$$M = N_\mathrm{i}N_\mathrm{h} + N_\mathrm{h} + N_\mathrm{h}N_\mathrm{o} + N_\mathrm{o} \tag{73}$$

memory values or *trainable parameters*, which are iteratively adjusted during the training process such that the network can fulfill a specified purpose (see Section 7.4). The network shown in Fig. 45 clearly has $M = 16$ trainable parameters. The number of neurons and interconnections serve as an indication of the power or computation capacity of a neural network. The average adult human brain is estimated to have 86×10^9 neurons [100], far exceeding current state-of-the-art ANN technology [98].

ANNs have found multiple applications, e.g. function approximation (nonlinear regression), probability estimation, pattern recognition (classification), clustering (grouping data by similarity) and prediction [101]. Finding the appropriate ANN structure for any application involves selecting and tuning many design parameters which typically require an iterative and/or trial-and-error approach. Some of these aspects are considered next with a focus on function approximation, which will be used in the temperature reconstruction application.

7.2 Data selection and preprocessing

An ANN can generally be thought of as a structure to express a functional relationship between explanatory (independent or input) variables and response (dependent, output or target) variables, i.e. finding a mapping between two or more datasets especially where the relationship is unclear or difficult to express using ordinary mathematical equations.

The first step in choosing the ANN structure is to select the appropriate input/output datasets. Although the initial dataset choices may seem obvious (especially the target

dataset) and inherently defined by the problem, further consideration could reduce the training load, resulting in faster training and improved performance.

For example, it may be useful to transform an input data signal which would make the relationship with the target signal more obvious to the ANN. Instead of using input x, it may be more instructive to use a function of x such as $\log(x)$ or $\sin(x)$. Using both x and functions of x concurrently could also be helpful. The choice of such functions is not necessarily obvious and may require analysis of the target dataset.

Data values should also be scaled or normalised such that they fall into a standard range (e.g. unitary bipolar $[-1; 1]$) or so that they have a specified mean and variance (e.g. $\mu = 0$ and $\sigma^2 = 1$) [101]. A dataset can be made unitary bipolar by applying the following transformation to each data value x:

$$x_{\text{bipolar}} = 2\left[\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right] - 1 \tag{74}$$

where $x_{\min}$ and $x_{\max}$ are respectively the minimum and maximum values of the dataset. After ANN training, it may be necessary to transform the data values (especially the trained ANN output) back to their original ranges. This can be achieved by applying the inverse transformation

$$x = \left[\frac{x_{\text{bipolar}} + 1}{2}\right][x_{\max} - x_{\min}] + x_{\min} \tag{75}$$

with $x_{\min}$ and $x_{\max}$ identical to the values used in (74). These extreme values should therefore be calculated and stored for each dataset during preprocessing. Note that (74) and (75) have implications for prediction or extrapolation - the trained ANN or function approximator should only be used to perform interpolation, i.e. new input values should not exceed the training ranges. No guarantees can be made regarding accuracy or stability if extrapolation is attempted. Even ANN architectures that perform time-series prediction, are still functioning as interpolators [101].

7.3 ANN structure

The ANN structure is defined by *hyperparameters*, which describe the ANN outside the learning algorithm. These hyperparameters include the number of layers, nodes per layer, choices and parameters defining the transfer functions, and parameters regarding training discussed in Section 7.4.8.

The number of nodes within the input and output layers follow directly from input/output

selection discussed in Section 7.2 above. Choosing the number of hidden layers and neurons per layer is much more difficult and depends on e.g. the type of problem and data available. However, according to the universal approximation theorem, a single hidden layer is sufficient to approximate a wide range of functions [98]. One method of determining the number of hidden neurons is through a grid search approach, where a range of values is evaluated or the number of hidden neurons is increased, from say $N_\mathrm{h} = 1$, until performance is satisfactory.

Many transfer functions are used in ANNs and the choice depends on the application and layer in the network. Fig. 46 shows three examples, including the classical standard logistic sigmoid (log-sigmoid or logsig), the hyperbolic tangent (tanh) and the rectified linear unit (ReLU) [98, 101].

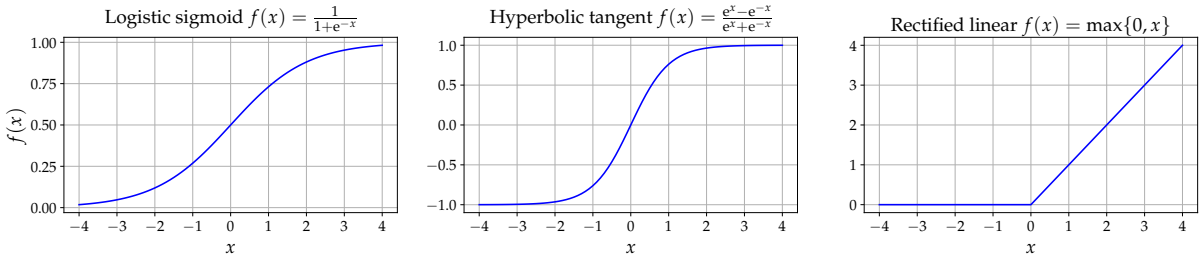


Figure 46: Three activation or transfer functions used in ANNs.

The choice of transfer function in combination with data scaling (discussed in Section 7.2 above) has implications for training and convergence speed. Transfer functions that are symmetric about the origin (e.g. $\tanh(x)$) often converge faster. Differences in gradient magnitudes and saturation points also have an effect on convergence [102]. The tanh transfer function is therefore typically preferred over the classical sigmoid function, although the ReLU has become the default recommendation in modern ANNs [98].

7.4 ANN training

Neural network training is an iterative process during which the weights are adjusted, with the aim of best approximating the relationship between input and output values or functions, given the available data and network structure. This section will explore several concepts required to understand ANN training.

7.4.1 Training, validation and testing data

After collecting the data to be analysed or processed using ANN techniques, the data are typically divided into the following three subsets.

- **Training set**: Data which are directly utilised to adjust the network weights.
- **Validation set**: Data used to monitor training progress, with the aim of stopping the process to prevent overfitting, which will result from excessive training.
- **Testing set**: Containing "unseen" data, or data not used in any way during training, which are used to predict the future performance of the network on new data.

The goal of training is to prepare the network for processing unseen data, based on the initial data discussed above. The training set usually contains the bulk (70–80%), whereas the other two sets each contain around 10–15% of the initial data. It is also possible to divide the data only into training and testing sets, and using other techniques to limit training (see Section 7.4.7). More data would then be available to perform training.

7.4.2 Training algorithm outline

Assume an example input set $\mathbf{x} = [x_1, x_2, \ldots, x_{N_\mathrm{i}}]$ is associated with the correct function or target value y_t. One of the simplest ANN training approaches to find the relation between $\mathbf{x}$ and y_t can then be summarised as follows.

1. Guess initial values for all the weights in the network, by assigning small randomly-distributed values. For example, uniformly distributed values between -0.5 and 0.5 can be chosen, if the inputs are normalised to fall within $[-1; 1]$ (see [101]).
2. Repeat the following steps for every training iteration.[19]
   2.1. Calculate the network output $y(\mathbf{x})$, for the given input set $\mathbf{x}$.
   2.2. Calculate the error between the calculated output and target value y_t, using the chosen loss or error function. Example loss functions include the difference $\varepsilon = y(\mathbf{x}) - y_\mathrm{t}$, squared difference ε^2, mean square error (MSE), etc.
   2.3. Propagate ε backwards through the network[20], and calculate the partial derivative $\frac{\partial \varepsilon}{\partial w_z}$ for every parameter w_z to obtain the gradient $\nabla\varepsilon$. The gradient measures how the error ε changes with respect to all weights in the network, which defines how much every parameter must be adjusted to reduce ε.

[19] See Appendix A for an example.

[20] This process is performed using the backpropagation algorithm, which is discussed in Appendix A.

2.4. Update each weight by applying the steepest or gradient descent rule [98]:

$$w_z^+ = w_z - \eta \frac{\partial \varepsilon}{\partial w_z} \tag{76}$$

or collectively for all weights

$$\mathbf{w}^+ = \mathbf{w} - \eta \nabla \varepsilon \tag{77}$$

where η is the learning rate or step size.

The learning rule in (76) is often simply expressed as the difference equation

$$\Delta w_z = w_z^+ - w_z = -\eta \frac{\partial \varepsilon}{\partial w_z} \tag{78}$$

indicating how much a given weight is adjusted.

7.4.3 Batch and incremental training

As indicated in the algorithm of Section 7.4.2 above, the same data values are typically used over and over to train the network. The training phase where all the values in the training set are used once is referred to as an *epoch*. There are three approaches of utilising the training set [98, 101]:

- **Batch training**, where the weights are updated once per epoch. All the training examples are processed by the network before the complete gradient is calculated and all the weights updated. For M input examples, M error values are therefore computed, from which the overall gradient is calculated. Examples include batch gradient descent and the Levenberg-Marquardt algorithm.
- **Incremental or online training**, where the weights are updated after every training example is processed. If the training set contains M examples, the network weights will therefore be updated M times for every epoch. The stochastic gradient descent (SGD) algorithm is a popular online training algorithm.
- **Minibatch training**, where the training set is divided into smaller subsets or minibatches. Each minibatch is then processed before updating the weights, which may be done using the average gradient of the individual examples in each minibatch.

Batch training provides a more accurate gradient estimate, but also requires more memory. Online training on the other hand can update the weights faster using single additional examples, but have a more noisy gradient signal. The minibatch approach combines the strengths of these two methods, and has therefore become the standard approach in many applications.

7.4.4 Numerical optimisation

ANN training is a numerical optimisation problem, where the parameters (weights $\mathbf{w}$) of an objective function[21] (error surface $\varepsilon(\mathbf{w})$) are to be found. An illustrative error surface is shown in Fig. 47.

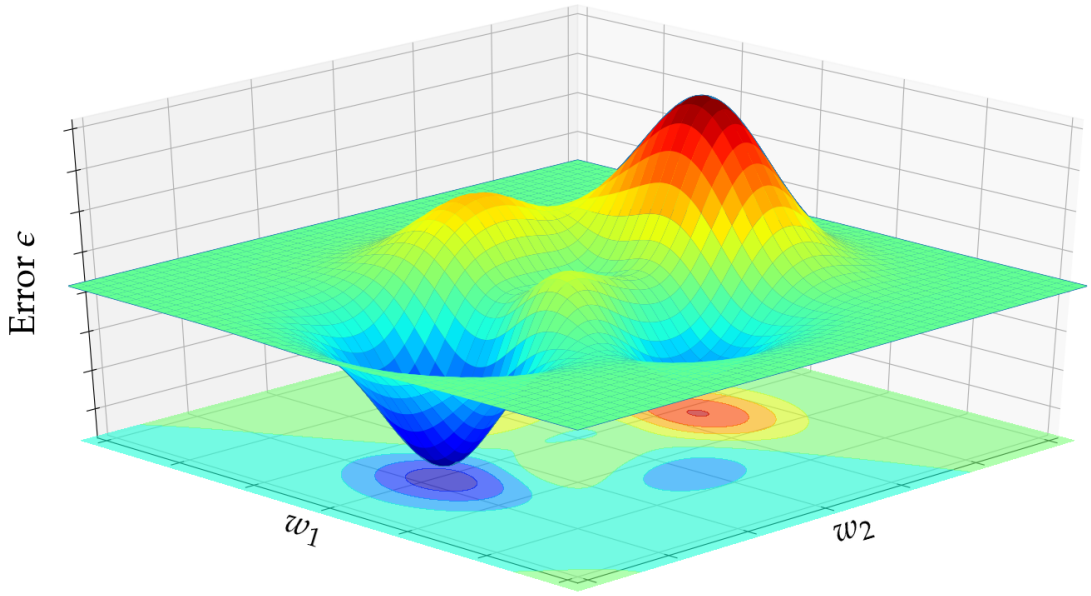


Figure 47: An illustrative error surface for a two-weight neural network.

Suppose a very simple network has only two weights, expressed by the vector $\mathbf{w} = [w_1, w_2]$. The error surface can then be depicted in 3D as shown in Fig. 47. Suppose initial values for the weights are randomly chosen as $\mathbf{w}^* = [w_1^*, w_2^*]$. The initial error would then be $\varepsilon(\mathbf{w})|_{\mathbf{w}=\mathbf{w}^*}$, which will decrease fastest if the next choice of weights is in the direction of the negative gradient of ε at $\mathbf{w}^*$. The gradient of the error can be written as [101]

$$\nabla\varepsilon(\mathbf{w}) = \begin{bmatrix} \frac{\partial}{\partial w_1}\varepsilon(\mathbf{w}) \\ \frac{\partial}{\partial w_2}\varepsilon(\mathbf{w}) \end{bmatrix} \tag{79}$$

where each gradient component (i.e. directional derivative) is used in the update rule of individual weights, which was given in (76).

By repeatedly applying (76) with sufficiently small step size η, the updated weight values will result in successively smaller error values. The error surface will thus be traversed in discrete steps, until either a local or global minimum is reached, at which point the steepest descent algorithm may become stuck.

[21] a function to be minimised/maximised within certain constraints

Traversing the error surface depends on the shape of the surface and the step size. Larger step sizes will typically progress faster towards a minimum, but convergence may become erratic, such that the optimisation algorithm may skip over, or jump in and out of minima. For a smaller step size, convergence is slower, but once a local minimum is reached, optimisation will stop as the algorithm doesn't have the ability to jump out of local minima in search of the global minimum. The maximum step size ensuring stable learning can be calculated as a function of maximum curvature of the error surface [101], although finding the optimal value for η would typically require experimentation.

The global minimum can also be found by training the network repetitively, each time starting from different initial conditions. It can be expected that for some initial conditions local minima will be reached, while for other the global minimum will be obtained. Selecting the weight values associated with the smallest error obtained during such experimentation, could be a viable approach.

7.4.5 More advanced learning rules

Several more advanced update rules based on (76) have been developed to improve the training process, principally to enhance convergence speed and the chances of finding a global minimum. Two examples of such learning rules are subsequently discussed.

Momentum

Learning can be enhanced by adding historical weight adjustments to the update rule. The weight adjustment of (78) can thus be modified to obtain the *momentum* rule

$$\Delta w_z(k) = -\eta \frac{\partial \varepsilon}{\partial w_z} + \alpha \Delta w_z(k-1) \tag{80}$$

with k the training iteration number and $\alpha \in [0, 1)$ the constant momentum parameter. The effect of including the momentum term is accelerated learning, ensuring stability with a larger step size [98, 101]. Gradients calculated in previous iterations continue to influence the current direction to continue moving along the historical path, resulting in smoothing out the trajectory and limiting the chances of getting stuck in local minima.

Adam

Learning can be enhanced further by using individual learning rates for each network weight, and allowing these learning rates to adapt throughout the course of training [98]. Adam (derived from "adaptive moment estimation") [103] is one such algorithm where the first two moments are used in the update rule, with exponentially-decaying contributions

from previous updates.

The first-order moment is a function of the previous moment and the gradient element associated with weight w_z:

$$m_z(k) = \beta_1 m_z(k-1) + (1-\beta_1)\frac{\partial \varepsilon}{\partial w_z} \tag{81}$$

at training iteration k, with decay rate $\beta_1 \in [0, 1)$. Similarly, the second-order moment is a function of the squared gradient element:

$$v_z(k) = \beta_2 v_z(k-1) + (1-\beta_2)\left(\frac{\partial \varepsilon}{\partial w_z}\right)^2 \tag{82}$$

with decay rate $\beta_2 \in [0, 1)$. The Adam update rule can then be written as

$$\Delta w_z(k) = -\eta \frac{\hat{m}_z(k)}{\sqrt{\hat{v}_z(k)} + \epsilon} \tag{83}$$

with η the learning rate, ϵ a small constant to ensure numerical stability, and bias-corrected moments

$$\hat{m}_z(k) = \frac{m_z(k)}{1-\beta_1^k} \text{ and } \hat{v}_z(k) = \frac{v_z(k)}{1-\beta_2^k} \tag{84}$$

Values for the constants are suggested as $\eta = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$ [103]. In (83) the learning rate η is constant, however by combining (83) and (84) the update rule can be expressed as

$$\Delta w_z(k) = -\eta \frac{\frac{m_z(k)}{1-\beta_1^k}}{\sqrt{\frac{v_z(k)}{1-\beta_2^k}} + \epsilon} \tag{85}$$

$$= -\eta \frac{\sqrt{1-\beta_2^k}}{1-\beta_1^k}\left(\frac{m_z(k)}{\sqrt{v_z(k)} + \sqrt{1-\beta_2^k}\epsilon}\right) \tag{86}$$

$$= -\eta(k)\frac{m_z(k)}{\sqrt{v_z(k)} + \hat{\epsilon}} \tag{87}$$

with the actual learning rate clearly changing as training progresses:

$$\eta(k) = \eta \frac{\sqrt{1-\beta_2^k}}{1-\beta_1^k} \tag{88}$$

and the modified stabilisation constant:

$$\hat{\epsilon} = \sqrt{1-\beta_2^k}\epsilon \tag{89}$$

The above equations show how individual weight values are updated. However, these equations will typically be implemented as vector operations in software, updating all weights in an ANN layer concurrently. Adam is the optimisation algorithm that will be used in the temperature reconstruction method presented in this report.

7.4.6 Function approximation example

An example will be considered in this section to illustrate some of the concepts considered thus far. Suppose samples are taken from the process described by the following function

$$g(t) = \underbrace{\sin\left(2\pi\frac{t}{T} + \frac{\pi}{5}\right)}_{\text{true signal}} + \overbrace{w(t)}^{\text{noise}} \tag{90}$$

with t the time variable, $T = 100$ s the signal period, and $w(t)$ a noise process. Samples are taken from $g(t)$ by increasing t from 0 to 100 s in discrete steps of $\Delta t = 1$ s.

An ANN approach is now needed to find the mapping between the following two datasets:

$$\begin{aligned} \text{Input}: &\quad \mathbf{t} = [0,\ 1,\ \ldots,\ 100] \\ \text{Output}: &\quad \mathbf{g} = [g(0),\ g(1),\ \ldots,\ g(100)] \end{aligned}$$

Note that the ANN has no knowledge of the true signal or noise characteristics; the two datasets $\mathbf{t}$ and $\mathbf{g}$ contain all available knowledge.

The standard ANN approach for function approximation (or regression) problems is using the universal approximator, which is a three-layer network with hyperbolic tangent hidden nodes and a linear output node [101]. This is the structure shown in Fig. 45, when the output transfer function is removed or changed from $f(x) = \tanh(x)$ to $f(x) = x$. The universal approximator architecture is also used in this example.

Following the guidelines presented in Section 7.2, the input and output datasets are each scaled using (74) so they are limited to $[-1; 1]$ before they are presented to the ANN. Furthermore, if $\mathbf{g}$ is displayed graphically, its sinusoidal structure will become clear. Two additional inputs (or helper functions) could therefore be chosen as $\sin(2\pi f_0 t)$ and $\cos(2\pi f_0 t)$, with f_0 the normalised frequency which can be estimated as the inverse of the number of samples within one period of the sine wave [15].

With reference to Fig. 45 and (74), the inputs and output of the ANN can then be assigned as follows.

$$x_1: \quad \mathbf{t}_{\text{bipolar}} \tag{91}$$

$$x_2: \quad \sin(2\pi f_0 t) \tag{92}$$

$$x_3: \quad \cos(2\pi f_0 t) \tag{93}$$

$$y(\mathbf{x}): \quad \mathbf{g}_{\text{bipolar}} \tag{94}$$

The SGD method with adaptive moment optimisation was chosen as training algorithm, and the weights of the ANN were initialised to uniformly-distributed random values within $[-0.5; 0.5]$. All data samples were used for training; no validation or testing were performed except for calculating MSE values after training. Training was performed over a fixed number of epochs, and no other mechanism (e.g. early stopping or regularisation) was used to limit adjustment of the weight values.

Four training scenarios were considered with results shown in Fig. 48. Each of the four figures display the true signal indicated in (90), the target dataset of (94), and the unique output obtained when applying the input data to the ANN after training is complete.

The true MSE shown in each case is the mean squared difference between the trained output and the true signal, whereas the data MSE is the mean squared difference between the trained output and the target data.

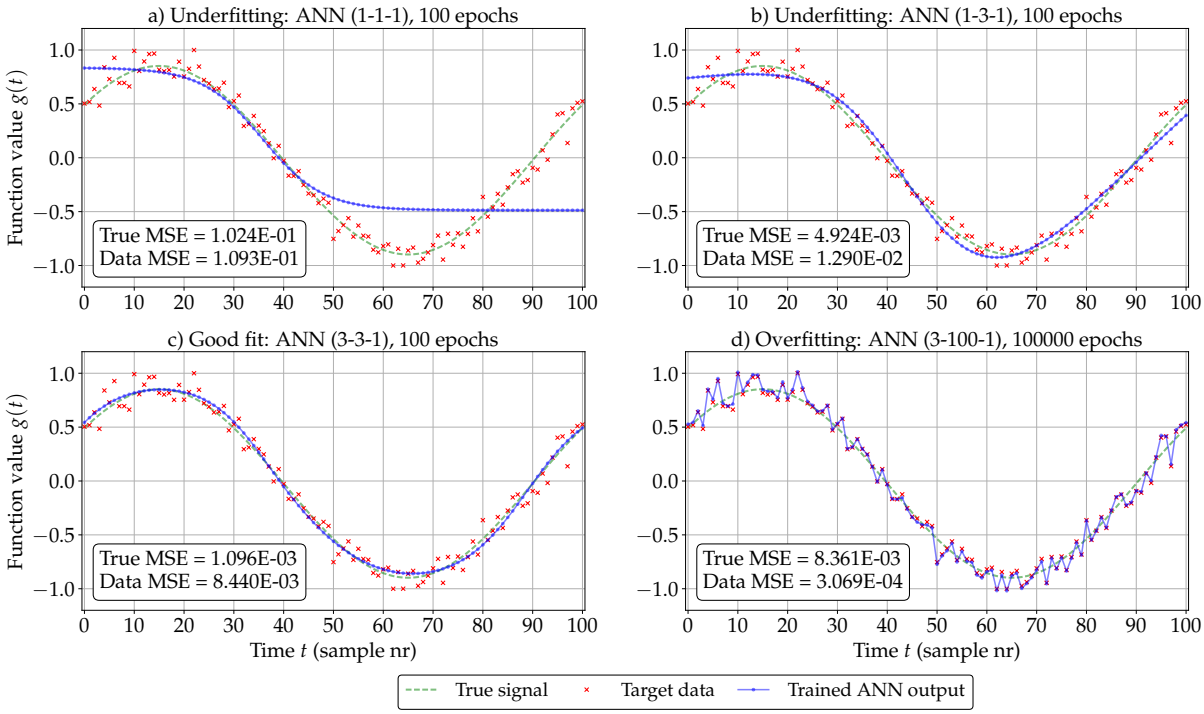


Figure 48: Trained ANN output for various scenarios.

In each case, either one input (x_1 in (91)) or three inputs (x_1 to x_3 in (91) to (93)) were applied to the network. In three cases (Fig. 48 a, b and c) training was performed over 100 epochs, and in one case (Fig. 48 d) training was performed for much longer to illustrate overfitting. The results shown in Fig. 48 a to d are summarised below.

a) With only x_1 as input and only 1 hidden neuron, the network is too small to capture the complexity of the target signal. The network therefore underfits, resulting in

the worst MSE values overall.

b) Still only using x_1 as input, but increasing the number of hidden neurons to 3, the network performance improves, although it still underfits.

c) Using 3 inputs (x_1, x_2 and x_3) with 3 hidden neurons, a good fit is obtained. The true MSE indicates this is the best performance of the four scenarios considered.

d) Using a large number of hidden neurons, and training for much longer, the ANN is able to find a mapping between the three inputs and the noisy target. This network is clearly too complicated and follows the noise in the target data. Although this scenario is an example of overfitting, it also serves as an illustration of the universal approximation theorem mentioned in Section 7.3.

7.4.7 Generalisation and regularisation

Neural network training should be monitored to ensure good performance on unseen data once training is done. As illustrated in the function approximation example of Section 7.4.6 above, excessive training poses the risk that the network will become fixated on the training set, with the possibility of even modeling noise. If training is continued unabated (i.e. weight adjustment is not limited or stopped at some point during training), the trained ANN will typically have superior performance on the training data, but poor performance on unseen data.

A mechanism to limit weight adjustment should therefore be built into the training algorithm to allow the trained ANN to generalise to scenarios not encountered during training. Training should thus be conducted to only extract enough information or general patterns from the training set (which should be representative of the larger target data set), before artefacts unique to the training set are embedded into the network memory, which will hamper network performance on unseen data.

A popular approach is to split the data into three sets as discussed in Section 7.4.1. Validation data are then used in-between training iterations to check whether the error continues to decrease while training continues. An increase in validation error is an indication that the network is beginning to overfit, or that its generalisation capability is starting to erode, and training should therefore stop. This strategy is known as "early stopping".

An alternative method is regularisation, where the error performance index or loss func-

tion is modified to include a term that penalises network complexity. Suppose the error performance is defined as

$$\varepsilon_{\mathrm{t}} = \frac{1}{2}\sum_{n=1}^{N_{\mathrm{o}}}(y_n - g_n)^2 \tag{95}$$

with N_{o} the number of ANN outputs, and y_n and g_n the n^{th} calculated and target outputs respectively. As discussed in Section 7.4.4, the training algorithm will attempt to minimise ε_{t} by choosing weight values that will minimise the difference $y_n - g_n$.

Regularisation can be implemented by modifying (95) to

$$\varepsilon_{\mathrm{t,reg}} = \varepsilon_{\mathrm{t}} + \frac{\lambda}{2}\sum_{m=1}^{M} w_m^2 \tag{96}$$

$$= \frac{1}{2}\sum_{n=1}^{N_{\mathrm{o}}}(y_n - g_n)^2 + \frac{\lambda}{2}\sum_{m=1}^{M} w_m^2 \tag{97}$$

where the second term is the weight penalty, which is $\lambda/2$ the sum-square of the M trainable parameters in the network. The regularisation parameter λ determines how much the weight penalty contributes towards the total error ε_{t}.

In addition to choosing weight values that will minimise the difference between output and target values, the optimisation algorithm is given the additional constraint that the weight values should also be small (controlled by the regularisation parameter λ). A more complex ANN tends to overfit and usually has weight values large in magnitude [101,104]. Using (97) instead of (95) will limit the growth of the network weights, limiting the possibility of overfitting.

It is therefore possible to obtain an ANN that is able to generalise better by simply limiting the weight magnitudes by introducing the regularisation term in the loss function. An additional advantage is regularisation does not require a validation set, leaving more data available for training.

The error derivative when using regularisation as in (96) can be expressed as

$$\frac{\partial \varepsilon_{\mathrm{t,reg}}}{\partial w_z} = \frac{\partial}{\partial w_z}\left\{\varepsilon_t + \frac{\lambda}{2}\sum_{m=1}^{M} w_m^2\right\} \tag{98}$$

$$= \frac{\partial \varepsilon_{\mathrm{t}}}{\partial w_z} + \frac{\partial}{\partial w_z}\left\{\frac{\lambda}{2}\left[w_1^2 + \ldots + w_z^2 + \ldots + w_M^2\right]\right\} \text{ with } 1 \leq z \leq M \tag{99}$$

$$= \frac{\partial \varepsilon_{\mathrm{t}}}{\partial w_z} + \lambda w_z \tag{100}$$

The effect of using the modified loss function given in (96) on the gradient is therefore the addition of λw_z, $1 \leq z \leq M$, to each gradient component [104].

As an example, the gradients given in (139) in Appendix A will be changed to

$$\nabla\varepsilon(\mathbf{w}) = \begin{bmatrix} \dfrac{\partial\varepsilon_t}{\partial w_{1,1}} + \lambda w_{1,1} & \dfrac{\partial\varepsilon_t}{\partial w_{1,2}} + \lambda w_{1,2} \\ \dfrac{\partial\varepsilon_t}{\partial w_{2,1}} + \lambda w_{2,1} & \dfrac{\partial\varepsilon_t}{\partial w_{2,2}} + \lambda w_{2,2} \end{bmatrix} \text{ and } \nabla\varepsilon(\mathbf{b}) = \begin{bmatrix} \dfrac{\partial\varepsilon_t}{\partial b_1} + \lambda b_1 \\ \dfrac{\partial\varepsilon_t}{\partial b_2} + \lambda b_2 \end{bmatrix} \tag{101}$$

if regularisation is implemented. Regularising the bias parameters $\mathbf{b}$ may affect performance negatively, in which case only the weight parameters $\mathbf{w}$ should be regularised (see p. 223 of [98]). Also, using unique values for λ for different layers in the network may also improve performance.

7.4.8 Training hyperparameters

Some hyperparameters describing the ANN structure were mentioned in Section 7.3. There are also hyperparameters that define how training should be performed (before commencing training) which may include, depending on the structure and algorithm used, the learning rate or step size η, the loss or error function ε, data segmentation (choosing the training, validation and testing set sizes), training batch size, weight initialisation, momentum parameter α, number of training iterations allowed, regularisation parameters, and the training algorithm itself.

7.5 Developing an AI solution

There are generally two approaches that can be followed, separately or in combination, when using AI to solve problems e.g. temperature reconstruction considered in this report:

1. **Using existing software**: There are many software packages and libraries available, e.g. TensorFlow, Theano, Caffe, Keras, Torch, Sci-kit Learn, etc. This approach is viable to experiment relatively quickly with many AI architectures and parameters once an appropriate software package has been identified. Although it is also possible to develop a complete solution using existing ML software, it is likely that part of the solution will need to be custom designed and coded.

2. **Developing a custom solution from first principles**: Enhanced performance (e.g. faster ANN training and evaluation) is possible when certain parts of the training algorithm is developed using lower-level coding (e.g. *C*). Performance can also be enhanced further using high-performance multi-processor computing platforms.

Both above-mentioned approaches were followed in this report. The TensorFlow [105] library was used in Python to develop first ML software versions to test and experiment a range of ANN structures and parameters. Enhanced training speed was further obtained by developing the core ML routines in *C* using Gnu Scientific Library (GSL) [106] as linear algebra library.

8 AI temperature reconstruction

Historically, temperature was measured at various locations that underwent changes as discussed in Sections 1.3 and 1.4. Weather stations were moved around, different measuring methods and instruments were used, and the environment surrounding weather stations changed over the course of history, as discussed in Section 2.

The resultant temperature record could be considered non-uniform, discontinuous, and sparse in view of the total number of locations and time frame over which recording took place. Reconstruction or remodelling is thus required to make sense of the data and to obtain a clear overall picture of historical temperature trends. In some cases however, weather stations remained in the same location and used the same equipment over extensive periods, enabling the study of long-term trends without remodeling data at these single sites.

In this section a method to reconstruct historical temperature data using ANN technology is introduced. The method can be used to estimate or infill missing data within single temperature series, and to estimate temperature series for geographical locations where no weather station ever existed. Area-weighted averages can then also be calculated to estimate trends over larger regions. The focus in this section is on monthly mean maximum temperature ($T_{\max}$) data, obtained from the BoM CDO archive [107].

8.1 Deniliquin case study

The map in Fig. 49 shows the area surrounding Deniliquin that will be used in this section to explain the temperature reconstruction technique conceptually. The map indicates 71 weather stations that contributed monthly mean $T_{\max}$ data to the CDO archive, some time from January 1910 to December 2018.

The 71 weather stations are listed in Table 3 and the twelve-month MA of each data series is shown in Fig. 50a[22]. The total number of monthly data samples available from all 71 records is shown in Fig. 50b for every month from Jan 1910 to Dec 2018.

[22] Beechworth Woolshed (ID 82137) is not shown in Fig. 50a, as this record does not contain at least one run of 12 consecutive months at any point in time, although this record contains 135 monthly values from Nov 1986 to Apr 2012.

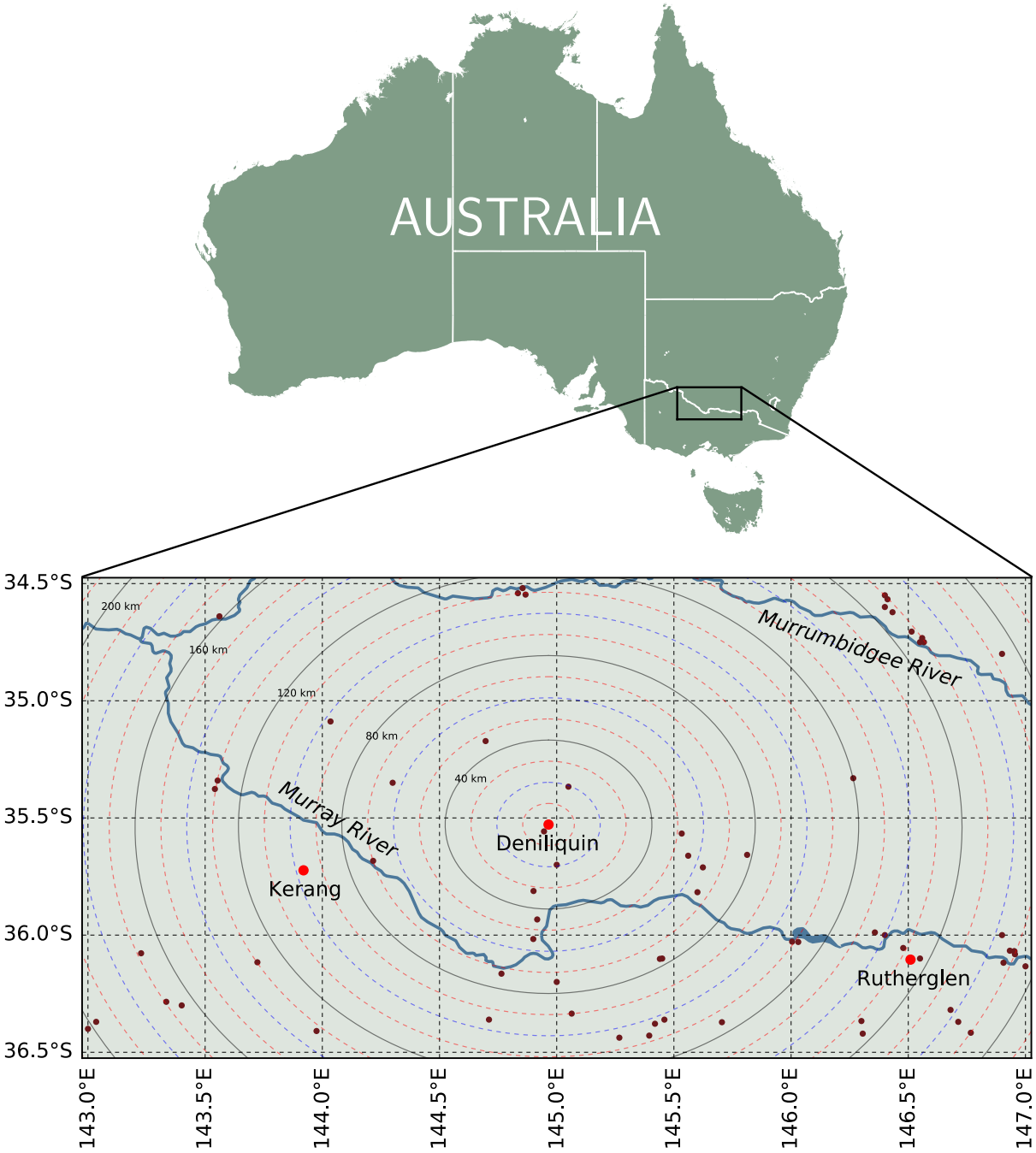


Figure 49: Area surrounding the town of Deniliquin (NSW) used as case study.

								Number of months		
#	ID	Station name	Lat (°S)	Lon (°E)	Alt (m)	Start	End	Present	Missing	Total
1.	80015	Echuca Aerodrome	36.165	144.764	99.8	Jan 1910	Dec 2018	1307	1	1308
2.	75031	Hay (Miller Street)	34.519	144.855	92.1	Jan 1910	Jan 2015	1260	1	1261
3.	82039	Rutherglen Research	36.105	146.509	174.3	Nov 1912	Dec 2018	1260	14	1274
4.	80023	Kerang	35.724	143.920	80.5	Jan 1910	Dec 2018	1247	61	1308
5.	49002	Balranald (RSL)	34.640	143.561	64.2	Jan 1910	Dec 2018	1191	117	1308
6.	74128	Deniliquin (V.I.C.)	35.528	144.965	95.7	Jan 1910	Jun 2003	1119	3	1122
7.	74034	Corowa Airport	35.989	146.357	141.4	Jan 1910	Dec 2018	1096	212	1308
8.	77042	Swan Hill Post Office	35.341	143.553	74.1	Jan 1910	Dec 1996	1044	0	1044
9.	82053	Wangaratta	36.367	146.300	150.0	Jan 1910	May 1987	911	18	929
10.	82001	Beechworth Composite	36.370	146.713	614.9	Jan 1910	Jun 1986	908	10	918
11.	80002	Boort	36.116	143.723	101.0	Jan 1910	Sep 1986	813	108	921
12.	74062	Leeton Caravan Park	34.567	146.411	139.8	Jan 1913	Dec 1975	739	17	756
13.	80043	Numurkah Post Office	36.100	145.450	112.5	Jan 1910	Jun 1977	702	108	810
14.	81049	Tatura Inst Sustainable Ag	36.438	145.267	113.0	Jan 1965	Dec 2018	648	0	648
15.	80006	Charlton Post Office	36.300	143.400	174.4	Jan 1910	Sep 1971	643	98	741
16.	74110	Urana Post Office	35.331	146.265	119.0	Jan 1914	Dec 1975	636	108	744
17.	80091	Kyabram	36.335	145.064	108.4	Jan 1965	Dec 2018	631	17	648
18.	74009	Berrigan Post Office	35.658	145.812	121.3	Jan 1910	Dec 1975	612	180	792
19.	74106	Tocumwal Airport	35.817	145.600	114.7	Dec 1970	Dec 2018	576	1	577
20.	78011	Donald Post Office	36.400	143.000	116.1	Jan 1910	Jan 1966	568	105	673
21.	74221	Narrandera Golf Club	34.733	146.559	179.2	Jan 1970	Sep 2014	524	13	537
22.	78072	Donald	36.370	143.035	116.4	Oct 1966	Oct 2000	409	0	409
23.	82138	Wangaratta Aero	36.420	146.306	151.0	May 1987	Dec 2018	379	1	380
24.	74039	Deniliquin Falkiner Memorial	35.367	145.050	92.7	Oct 1947	Nov 1977	360	2	362
25.	78042	Wycheproof	36.078	143.227	112.1	Jan 1910	Sep 1938	345	0	345
26.	82085	Rutherglen Viticulture College	36.100	146.550	190.4	Jan 1910	May 1937	329	0	329
27.	81084	Lemnos (Campbells Soup)	36.361	145.459	115.5	Jan 1965	Jan 1996	326	47	373
28.	72160	Albury Airport AWS	36.069	146.951	166.3	May 1993	Dec 2018	308	0	308
29.	81124	Yarrawonga	36.029	146.030	128.5	May 1993	Dec 2018	306	2	308
30.	80049	Rochester	36.361	144.711	116.1	Dec 1940	Jul 1975	298	118	416
31.	72146	Albury Airport	36.069	146.953	162.8	Jun 1983	Nov 2007	293	1	294
32.	74148	Narrandera Airport AWS	34.705	146.514	142.3	Nov 1970	Dec 2018	274	304	578
33.	81125	Shepparton Airport	36.429	145.395	114.0	Jul 1996	Dec 2018	270	0	270
34.	77094	Swan Hill Aerodrome	35.377	143.542	70.6	Dec 1996	Dec 2018	265	0	265
35.	74258	Deniliquin Airport AWS	35.557	144.946	95.8	Jun 1997	Dec 2018	259	0	259
36.	74069	Mathoura State Forest	35.812	144.901	111.5	Mar 1949	Dec 1969	249	1	250
37.	74037	Yanco Agricultural Institute	34.622	146.433	164.6	Aug 1999	Dec 2018	233	0	233
38.	74076	Moira Pastoral Co.	35.933	144.917	97.2	Apr 1938	Dec 1956	219	6	225
39.	82100	Bonegilla	36.133	147.000	200.3	Jun 1968	Jun 1986	217	0	217
40.	75038	Koondrook State Forest	35.683	144.217	89.9	Feb 1940	Dec 1956	203	0	203
41.	72097	Albury Pumping Station	36.082	146.955	166.3	Jun 1970	Aug 1986	195	0	195
42.	75083	Werai	35.350	144.300	76.7	Jan 1910	Jan 1925	180	1	181
43.	80128	Charlton	36.285	143.334	131.8	Jul 2004	Dec 2018	174	0	174
44.	82056	Wodonga	36.118	146.906	158.2	Mar 1954	May 1968	170	1	171
45.	74051	Gulpa Island	35.700	145.000	97.5	Apr 1938	Nov 1951	163	1	164
46.	82038	Rutherglen Post Office	36.055	146.478	159.6	Jan 1910	Jun 1925	161	25	186
47.	75080	Wanganella (Zara)	35.172	144.697	88.2	Jan 1914	Mar 1927	159	0	159
48.	74072	Matong Forestry	34.800	146.900	172.1	Nov 1941	Dec 1956	153	29	182
49.	74133	Yanco Exp. Farm	34.600	146.400	137.8	Apr 1922	Sep 1935	151	11	162
50.	80101	Numurkah	36.102	145.440	108.2	Jul 1977	Apr 1992	149	29	178
51.	75019	Hay Airport AWS	34.541	144.834	90.3	Jun 2007	Dec 2018	139	0	139
52.	82137	Beechworth Woolshed	36.319	146.680	333.9	Nov 1986	Apr 2012	135	171	306
53.	75175	Hay CSIRO AWS	34.547	144.867	91.7	Apr 1995	Sep 2007	131	19	150
54.	81013	Dookie Agricultural College	36.372	145.705	188.5	Jan 1965	Oct 1975	130	0	130
55.	74134	Jindera (Murrumbateman)	36.000	146.900	312.3	Jan 1914	Nov 1925	130	13	143
56.	81057	Yarrawonga Post Office	36.028	146.004	129.3	Jan 1965	Sep 1975	129	0	129
57.	74063	Leeton Rice Research	34.550	146.400	140.1	Jan 1941	May 1951	97	28	125
58.	82050	Wahgunyah Nursery	36.000	146.400	134.4	Jan 1925	Oct 1932	91	3	94
59.	81044	Goulburn River @ Shepparton	36.378	145.419	114.3	Jan 1965	Jun 1972	90	0	90
60.	72059	Albury Grammar School	36.067	146.933	171.8	Mar 1962	Nov 1969	85	8	93
61.	80092	Wyuna	36.200	145.000	101.0	Jan 1913	Dec 1918	72	0	72
62.	82059	Stanley State Forest	36.417	146.767	745.5	Nov 1952	Jun 1966	68	96	164
63.	75046	Moulamein Post Office	35.089	144.035	73.7	Aug 1970	Dec 1975	65	0	65
64.	74077	Moira State Forest	36.017	144.900	115.6	Jan 1965	Dec 1969	59	1	60
65.	74083	Narrandera State Forest	34.750	146.567	178.9	Jan 1965	Dec 1969	58	2	60
66.	74253	Finley (CSIRO)	35.711	145.624	110.3	Apr 1995	Mar 2001	54	18	72
67.	74210	Deniliquin Aero	35.533	144.967	95.2	Jan 1942	Sep 1945	44	1	45
68.	74093	Finley (Strathdrummond)	35.567	145.533	107.9	Jan 1965	Sep 1968	44	1	45
69.	74082	Narrandera Post Office	34.750	146.550	149.6	Jan 1962	Feb 1970	39	59	98
70.	74023	Finley Airport AWS	35.661	145.561	110.6	Mar 2001	Sep 2003	31	0	31
71.	80099	Serpentine Loddon Valley H'Way	36.409	143.975	119.9	Apr 1969	Sep 1971	29	1	30
							Total:	27,352	2,192	29,544

Table 3: Station record details for all 71 locations in the Deniliquin area.

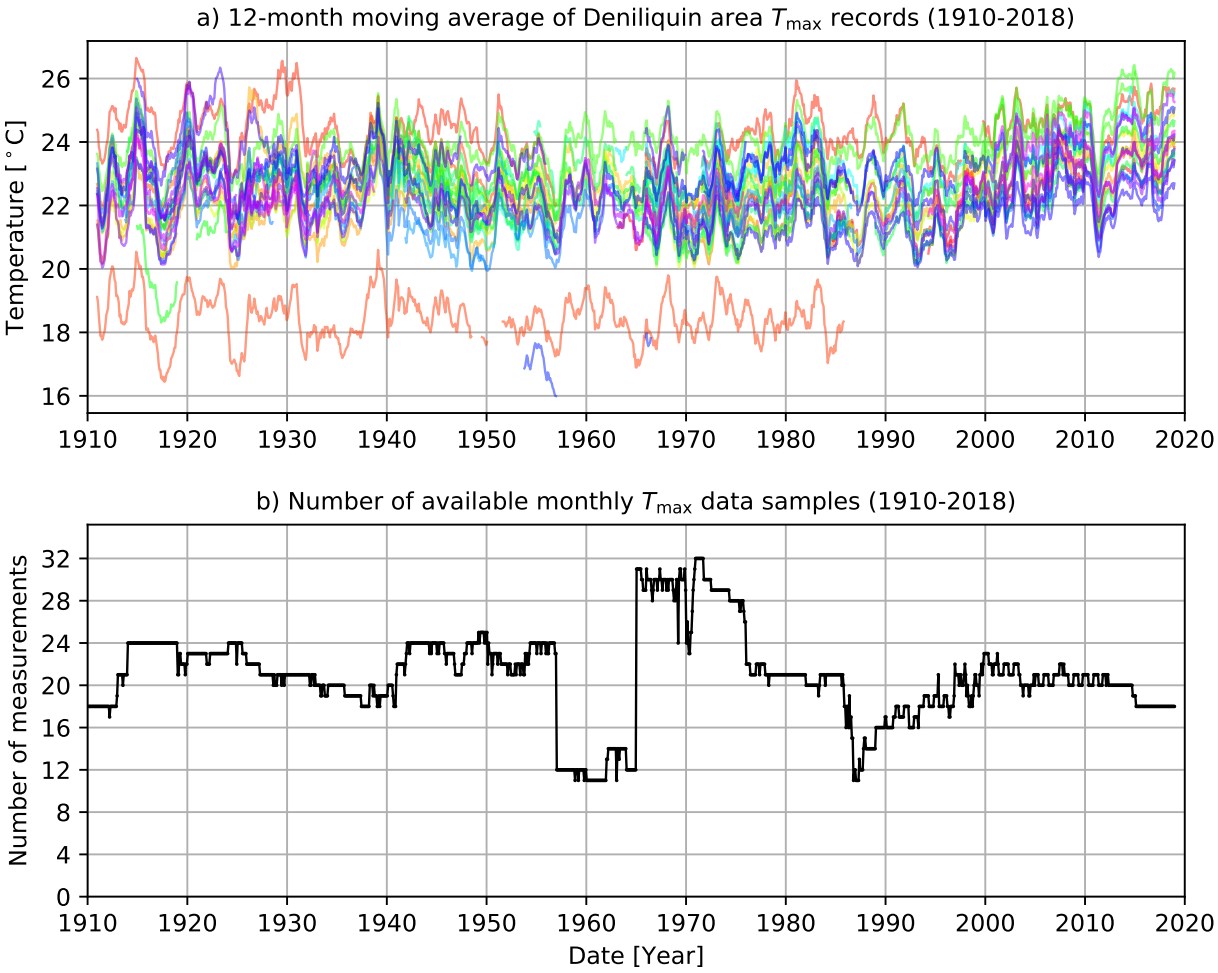


Figure 50: Monthly mean maximum temperature series of all 71 locations.

8.1.1 Infilling existing time series

There are 27,352 monthly samples available from the 71 weather stations shown in Fig. 49, over the 109-year period 1910–2018. The overall record can therefore said to be

$$\frac{100 \times 27,352}{109 \times 12 \times 71} = \frac{2,735,200}{92,868} = 29.453\% \tag{102}$$

complete.

The monthly T_{max} record of Deniliquin (Visitor Information Centre ID 74128, located in the town centre) is shown in Fig. 51. This station was active from Feb 1858 to Jun 2003, although only the data starting from Jan 1910 are considered in line with the official ACORN-SAT reconstruction methodology (see Sections 2 and 3.2). Another Deniliquin weather station (ID 74258) was opened in Jun 1997 at the airport, which effectively replaced the town centre station after both stations were active in parallel for six years.

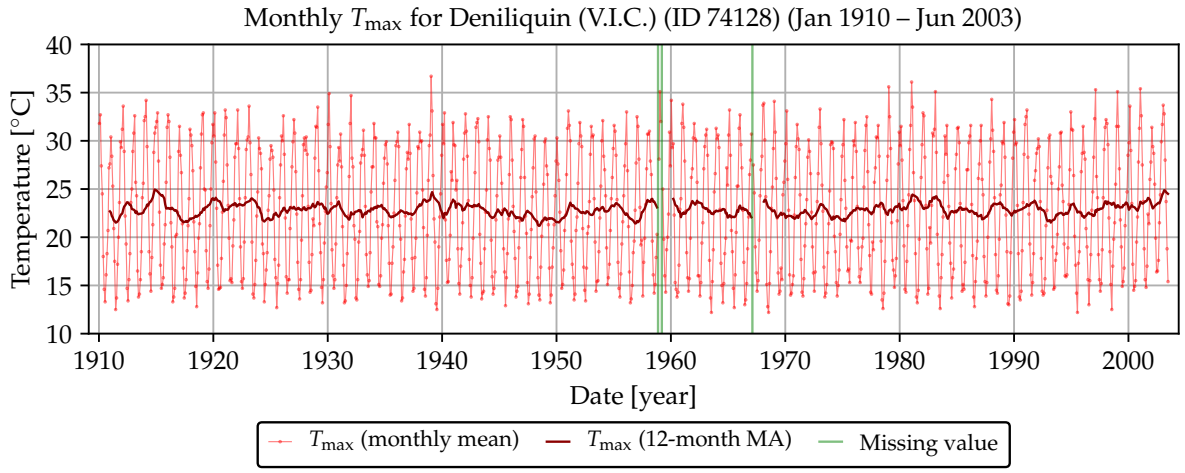


Figure 51: Monthly mean maximum temperature series of Deniliquin (74128).

As shown in Fig. 51, there are only 3 missing values in the monthly T_{max} record of Deniliquin (74128) over its active period. These missing values can be recovered from neighbouring stations, and data beyond the active period can also be estimated from neighbouring stations. A complete record over Jan 1910 to Dec 2018 can thus be formed for every station in the study area.

Techniques which can be used to infill or estimate missing data in existing temperature series will subsequently be considered. The performance of these techniques will be evaluated and compared using the existing values of the Deniliquin (74128) record.

8.1.2 Estimating time series at unsampled sites

The same techniques used to infill missing values within existing time series can be extended to estimate temperature data at arbitrary locations, where no actual measurements were ever taken. This involves spatial interpolation techniques, as discussed in Section 4.

To estimate a historical temperature profile for a region, virtual weather stations can be created on a regular grid, and the temperature series for each location can then be estimated using existing weather data for the region. An effectively area-weighted average trend over time can then be determined.

An example grid for the Deniliquin area is shown in Fig. 52, where virtual weather stations are spaced at 0.125° over both latitude and longitude. A total of $17 \times 33 = 561$ locations to be estimated is thus created.

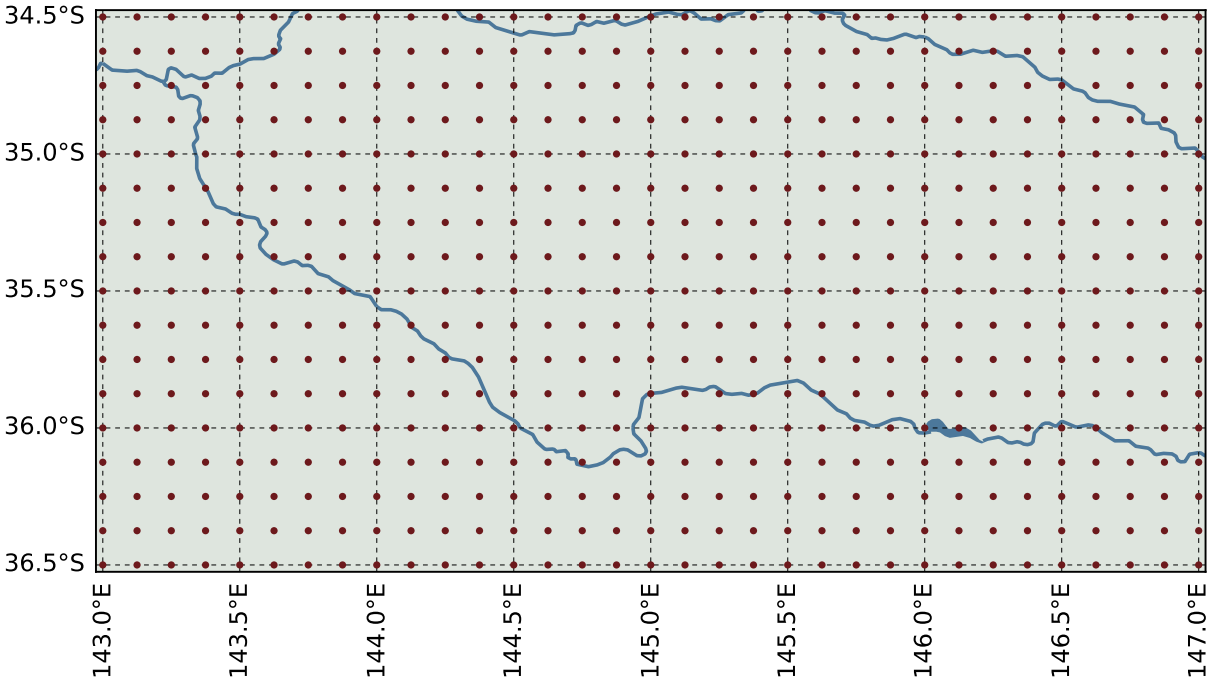


Figure 52: Map containing 561 virtual weather stations (spaced at 0.125° in both dimensions) forming a regular grid.

8.2 Estimation techniques

The temperature estimation technique based on AI is introduced here, with the IDW technique (see Section 4.1) used as benchmark.

8.2.1 Neural network temperature estimation

The ANN estimation technique evaluated in this study uses the structure shown in Fig. 53 below. To estimate a given missing monthly value, the network is trained on data from a selection of neighbouring weather stations. The inputs are the location (latitude and longitude) of a given neighbour, and the output is the $T_{\max}$ value for that location.

As discussed in Sections 7.2 to 7.4, the inputs and output are scaled to fit the interval $[-1; 1]$ and the hidden layer contains hyperbolic tangent transfer functions. The training algorithm is SGD with adaptive moments, and regularisation is used to prevent overfitting.

The ANN shown in Fig. 53 has a 2-3-1 structure, although the number of hidden nodes was varied over a range to find the best performance.

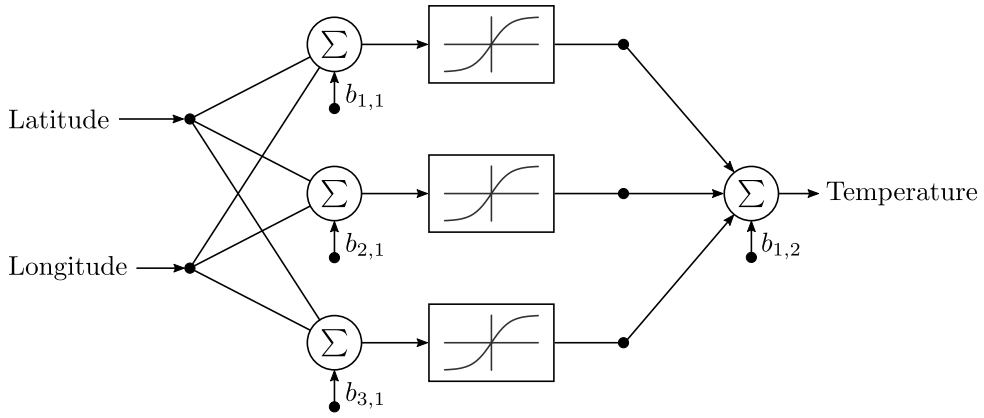


Figure 53: Neural network structure to estimate temperature data at a given location.

8.2.2 Inverse distance weighting

As discussed in Section 4, the temperature T at location x_0 can be estimated as the weighted average of N neighbouring temperature values, using the formula

$$\text{Plain IDW:} \quad \hat{T}(x_0) = \sum_{n=1}^{N} \lambda_n T(x_n) \tag{103}$$

with λ_n the weight value associated with the temperature value measured at location x_n. Using the IDW method, the weights can be calculated using (2) repeated here as

$$\lambda_n = \frac{d_n^{-p}}{\sum_{i=1}^{N} d_i^{-p}} \tag{104}$$

with d the distance between location x_0 and x_n, and p the distance-power parameter[23]. The choice of p will influence the performance, and the optimal value can be determined experimentally by evaluating a range of values.

For the infilling of existing time series, the performance of IDW can be improved by taking the relative offsets between temperature values at different locations into account. This can be achieved by modifying (103) to

$$\text{Modified IDW:} \quad \hat{T}(x_0) = \sum_{n=1}^{N} \lambda_n \left[T(x_n) + \Delta T_\mu(x_0, x_n)\right] \tag{105}$$

with the mean difference between available temperature values at locations x_0 and x_n

[23] As mentioned in Section 4.1, p controls how much emphasis is given to nearby locations. When $p = 0$, all neighbours are weighted equally such that $\hat{T}$ becomes the mean of all neighbouring values. As p increases, closer locations contribute more to $\hat{T}$.

given by

$$\Delta T_{\mu}(x_0, x_n) = \frac{\sum_{k=1}^{N_o} T(x_0, k) - T(x_n, k)}{N_o} \tag{106}$$

with $T(x_0, k)$ and $T(x_n, k)$ the k^{th} pair of valid (i.e. not missing data) temperature samples, and N_o the total number of valid or overlapping pairs. Note that N_o may be unique to every combination of location pairs. For example, location x_0 and x_1 may have 20 overlapping samples, while x_0 and x_2 may only have 10 overlapping samples.

8.3 Performance evaluation

The performance of an interpolation technique can be evaluated by removing one or more known values from the time series that is to be interpolated, and then estimating the removed value/s again. The difference (or error) between the known and estimated values can then serve as an indication of how accurate the technique would be when actual missing values are to be estimated.

This approach can be used to evaluate both the infilling of existing time series (see Section 8.1.1) and the estimation of time series at unsampled locations (see Section 8.1.2). The performance of infilling single values in existing series can be determined through LOOCV (see Section 6.3), which can be used to estimate performance over a whole series by removing and estimating every sample in the series in turn. The average performance can then be expressed using the MAE score. The concept can be expanded by leaving out and estimating more samples, up to leaving out all samples and estimating the whole time series.

When the whole time series is left out (in other words, known samples at a given location are not utilised in the estimation process) the MAE would indicate the performance of the technique to estimate data values associated with unsampled locations. Using this approach, the temperature data of a known weather station located at x_0 are estimated using its neighbours, as if no data were collected at x_0. The MAE is then calculated as the mean absolute difference between the estimated and true temperature series.

This section considers the evaluation of the IDW and ANN estimation techniques using the LOOCV approach for infilling single values. The T_{max} record of Deniliquin (74128) shown in Fig. 51 is used as case study. The entire T_{max} record of Deniliquin is also estimated from neighbouring data without using the Deniliquin data, to indicate the performance

of estimating unsampled locations.

8.3.1 Performance of plain and modified IDW

The results of using plain IDW (i.e. applying (103) with (104)) to estimate the Deniliquin (74128) record of Fig. 51 are shown in Fig. 54. Both the original raw CDO series T_{CDO} and the IDW-estimated series $\hat{T}_{\text{IDW}}$ are shown in the top graph. The absolute difference between these two series is shown in the bottom graph with the MAE. The three missing values in the raw series are also indicated.[24]

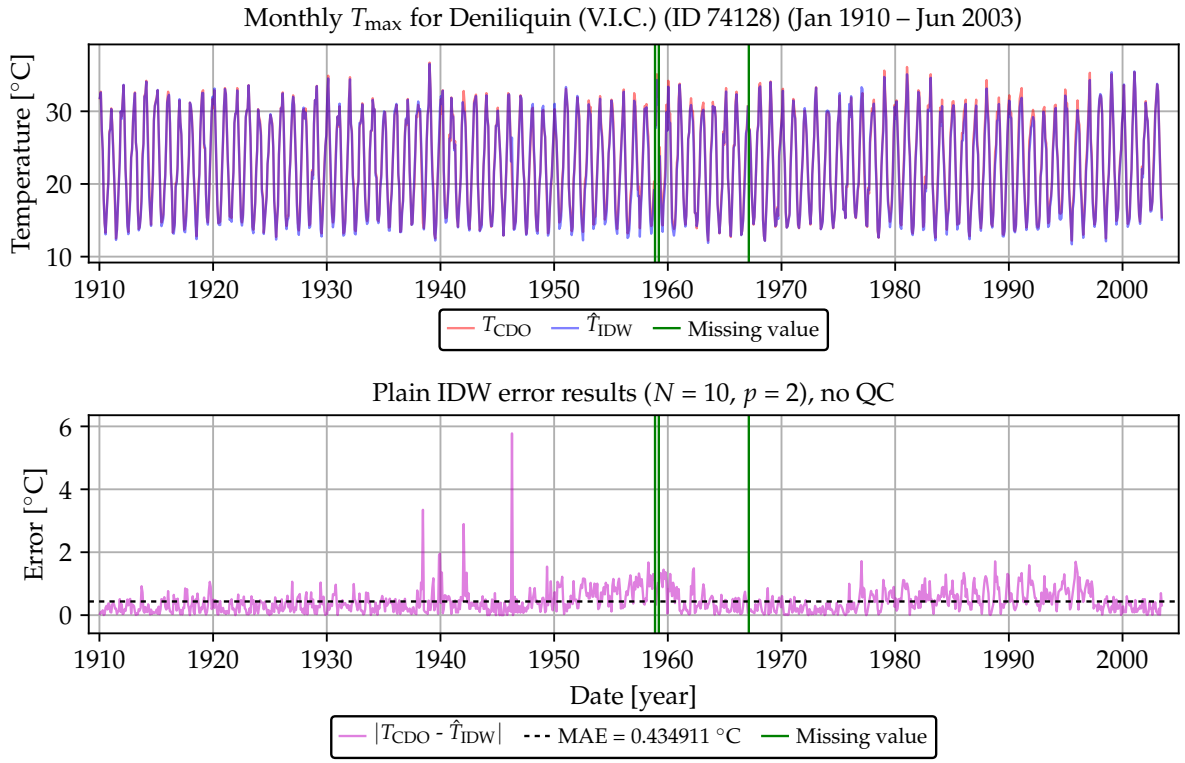


Figure 54: Estimation results for Deniliquin using plain IDW.

Fig. 54 therefore shows the results when Deniliquin (74128) T_{max} is estimated from its neighbours (in this case the closest 10 neighbours for any given month) without using any data from the series itself. If data from the series are used (after removing the single value under test), the performance can be improved as shown in Fig. 55, where the modified estimation formula of (105) was applied.

[24] As previously mentioned, the aim here is not to estimate the actual missing values, but to exclude known values and estimate them again to indicate how accurate the actual missing values could be estimated.

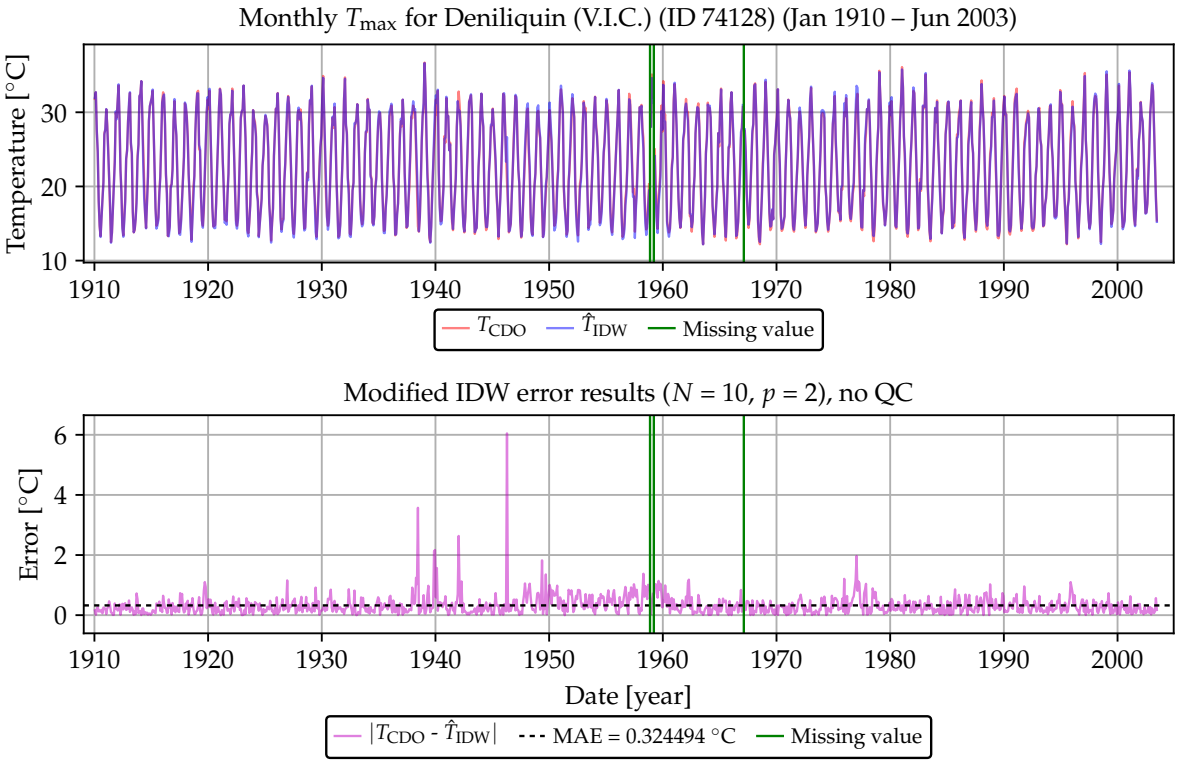


Figure 55: Estimation results for Deniliquin using the modified IDW method.

As shown in Figs. 54 and 55, plain IDW achieved an average error of 0.434911 °C, whereas modified IDW achieved an error of 0.324494 °C. The average error has therefore been reduced using the modified IDW technique, although the large error spikes have not been removed.

The results shown in both Figs. 54 and 55 have been obtained using the raw T_{max} data of Deniliquin (74128) without performing any QC first. Without QC, data errors are propagated to neighbouring stations when infilling techniques are used, causing large estimation errors shown as spikes in the two figures.

8.3.2 Quality control performance gains for IDW

QC can be performed on a single monthly temperature value T_{test} using the nearest-neighbour techniques presented in Section 5.2. Two QC variants, summarised below, will be considered.

- **Raw QC**: A data array containing T_{test} and its closest N_g neighbouring values for the same month is created. If T_{test} exceeds the 3σ-CI, it is regarded as an outlier and removed from the temperature record (see Section 5.2.2 for more details).

- **Adjusted QC**: Using the same approach as above, except that each data value in the array is adjusted according to the difference between the mean values of the central and neighbouring station records. More details are available in Section 5.2.2 with an illustration in Fig. 23.

By first performing QC (using the adjusted QC method with $N_g = 20$ neighbours and requiring at least $N_o = 6$ months of overlap between each central and neighbouring station pair, as discussed in Section 6.1.2) and then applying the modified IDW technique, the results of Fig. 56 are obtained. The largest error of approximately 6 °C at April 1946 shown in Figs. 54 and 55 has clearly been removed through QC. The value causing the large error originates from the Gulpa Island record, as illustrated in Figs. 21 and 24.

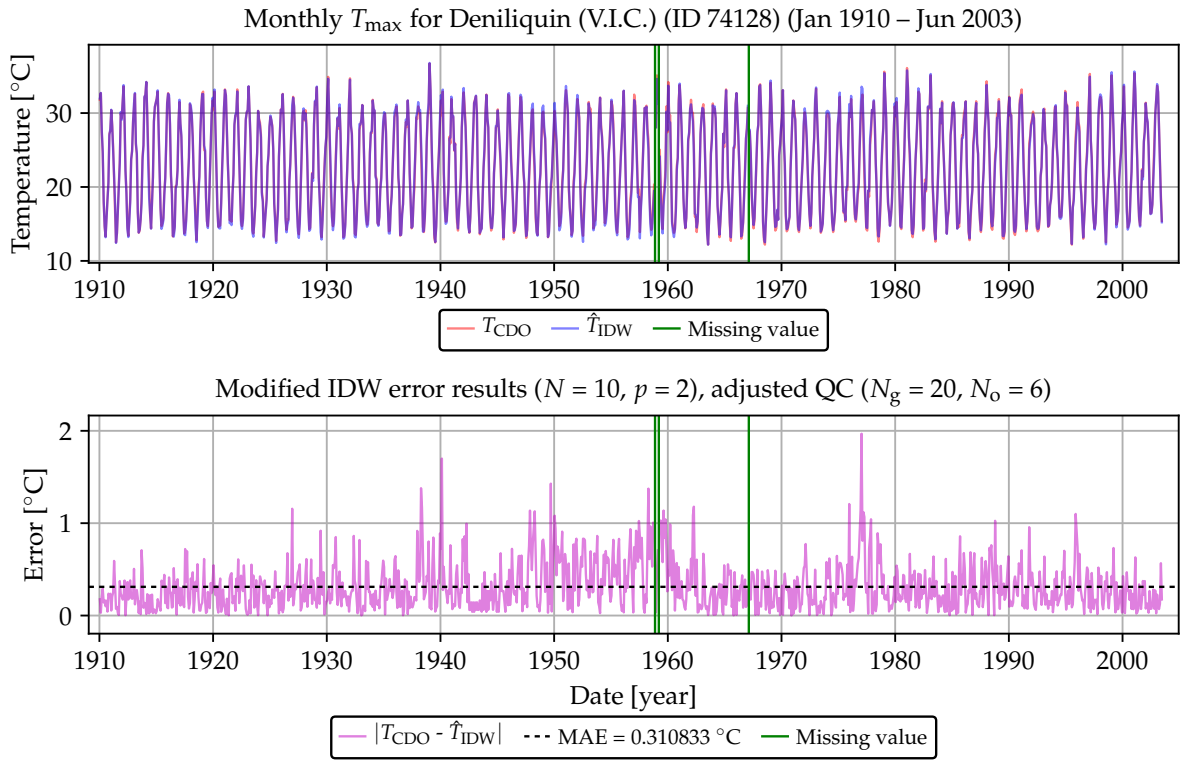


Figure 56: Estimation results for Deniliquin using the modified IDW method with QC.

Table 4 shows MAE values for various QC parameters, when plain and modified IDW are used to estimate missing values of the monthly T_{max} record of Deniliquin (74128). In plain IDW, the entire Deniliquin record is left out when each monthly value is estimated. In modified IDW all available data are utilised to estimate every monthly value; the modified IDW values are therefore LOOCV results.

Table 4 illustrates that modified IDW is better than plain IDW, and the adjusted QC method has the same or better performance than the raw QC method. If the number of

neighbours N_g used to perform QC is too small, QC has no effect and the MAE results are identical to the case where no QC has been attempted. The number of minimum required overlapping months N_o has a negligible effect, as the minimum requirement are exceeded in most cases. The best performance of the IDW scenarios shown is clearly for $N_g = 20$, as worse performance is achieved when too many neighbours are used.

QC par		Plain IDW		Modified IDW		Plain ANN		Modified ANN	
N_g	N_o	Raw QC	Adjusted QC	Raw QC	Adjusted QC	Raw QC	Adjusted QC	Raw QC	Adjusted QC
10	1	0.434911	0.434911	0.324494	0.324494	0.313522	0.313522	0.283516	0.283516
10	6	0.434911	0.434911	0.324494	0.324494	0.313522	0.313522	0.283516	0.283516
20	1	0.426176	0.420217	0.315870	0.310833	0.311131	0.309531	0.281116	0.279972
20	6	0.426176	0.420217	0.315870	0.310833	0.311131	0.309531	0.281116	0.279972
30	1	0.427150	0.421973	0.317220	0.311278	0.313006	0.308613	0.282849	0.279229
30	6	0.427150	0.421918	0.317220	0.311268	0.313006	0.308442	0.282849	0.279272
No QC		0.434911		0.324494		0.313522		0.283516	

Table 4: MAE results for IDW ($N = 10$, $p = 2$) and ANN for Deniliquin (74128) $T_{\max}$.

8.3.3 Optimal IDW parameters

Given a fixed geographical scenario with a central location where spatial interpolation is required, the performance of IDW depends on i) the number of neighbours utilised, and ii) how much influence is given to each neighbour in terms of its distance from the central location.[25]

From (103) and (104) these two factors are controlled by the N and p parameters. The optimal values for N and p can be determined using a grid search, with an example shown in Fig. 57 below. The optimal parameters for plain IDW without performing QC are shown to be $(N, p) = (10, 2)$, which were used in the analysis above.

All plain IDW results shown in Table 4 are optimal; i.e. no other (N, p) parameters will result in smaller MAE values. However, the modified IDW can be improved slightly. For $N_g = 20$, the raw QC result can be improved from 0.315870 to 0.294313 ($N = 9, p = 0$), and the adjusted QC result can be improved from 0.310833 to 0.290585 ($N = 9, p = 1$).

8.3.4 Performance of ANN

The ANN structure shown in Fig. 53 can be used to estimate the temperature at a given location (specified by latitude and longitude) after the ANN has been trained using the

[25] Assuming the closest N neighbours from the central location is included to perform interpolation. The performance could be improved by using only a subset of neighbours, based on e.g. neighbour data quality, or correlation with the central location data.

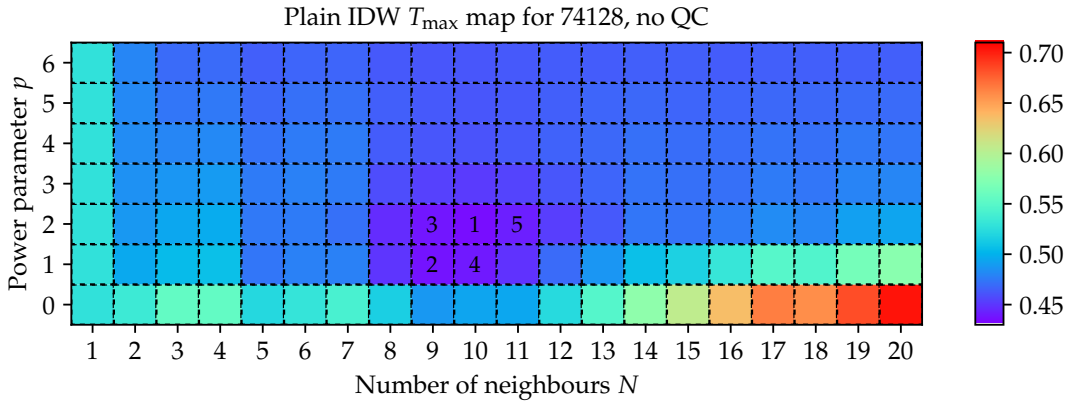


Figure 57: Heatmap grid search results to find optimal (N, p) parameters for Deniliquin.

temperature data at surrounding locations. Using the exact same data as for plain IDW, but using the ANN method instead, the results of Fig. 58 are obtained.

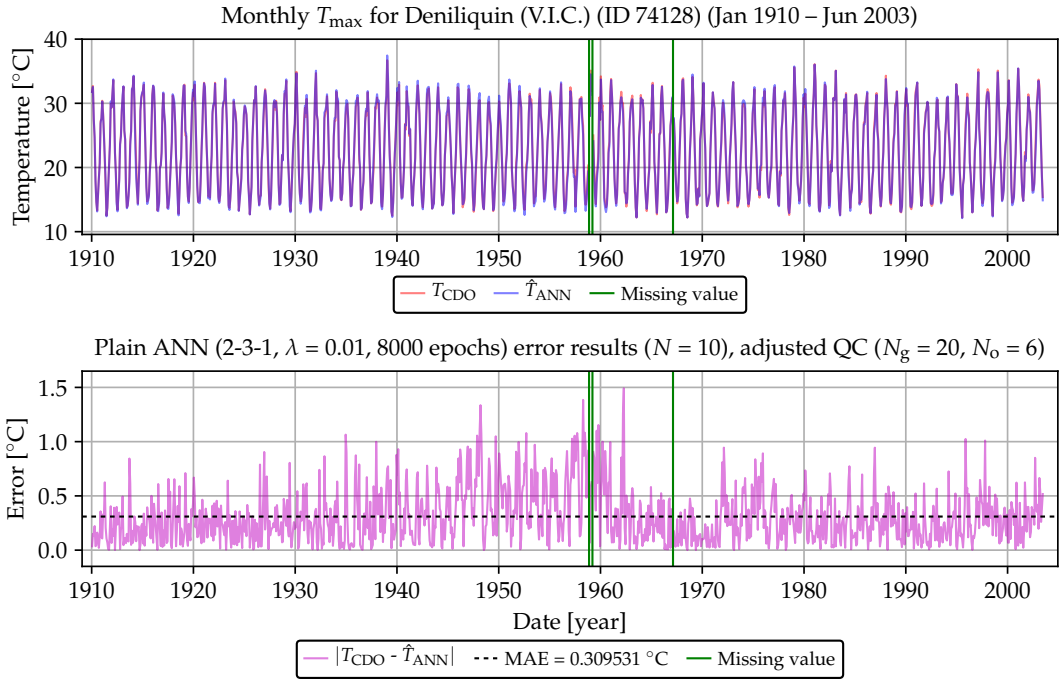


Figure 58: Estimation results for Deniliquin using plain ANN.

More ANN error results are also provided in Table 4, indicating that the ANN method is superior to both IDW methods for this case study. Note that the plain ANN method achieves a smaller error than the modified IDW method, although modified IDW uses more data (data of the record to be infilled). By adapting the ANN method (by modifying the data using the mean difference between data series as in (105)), improved results than those obtained by plain ANN are obtained as shown in Table 4.

The plain ANN method is summarised as follows. For every available monthly value

in the Deniliquin (74128) record, the 10 closest neighbouring weather stations (with an available value for the same month) are identified and used to train the ANN. The latitude and longitude of each neighbour are used as inputs to the ANN, with the raw monthly temperature value as the output (or target value). The Deniliquin $T_{\max}$ value is then estimated using its latitude and longitude, and compared with the recorded value.

8.3.5 Comparative results

Fig. 59 shows the MAE performance of the estimation techniques presented in Section 8.2 applied to the $T_{\max}$ data of the longest 8 stations listed in Table 3, after performing adjusted QC.

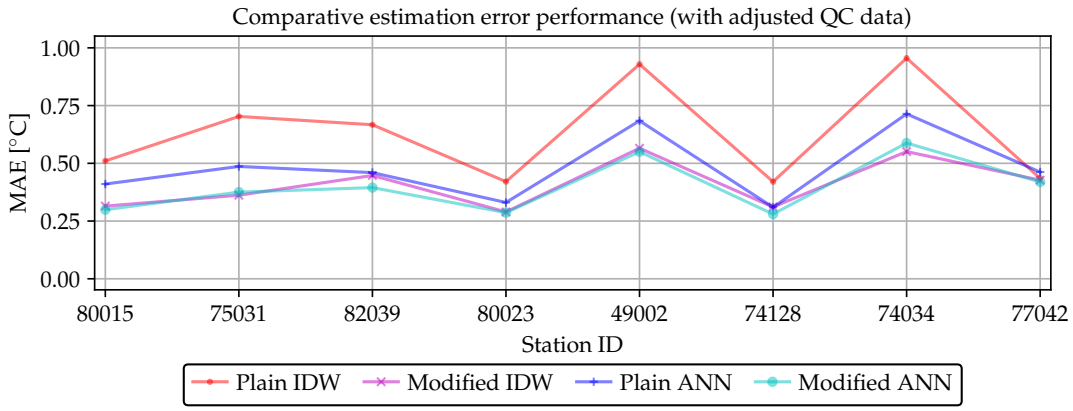


Figure 59: MAE results for the three estimation techniques.

The two plain techniques can be used to estimate temperature data at unsampled locations, whereas the modified techniques are intended to infill missing data in existing time series. The plain techniques can also be used to infill missing data, although this is not what they were designed for. The performance of each plain technique will be the same whether it is used to estimate data at unsampled locations or if it is used to infill missing data, as the plain techniques assume there are no data available at the location where estimation is to be performed.

From Fig. 59 it is clear that modifying the data improves performance for all cases shown. Plain ANN is better than plain IDW in almost all cases shown, and the ANN technique would therefore generally be superior in estimating data at unsampled locations. The two modified techniques show similar results, and would therefore provide similar infilling performance.

9 Discussion

This report considered the reconstruction of historical temperature data with a focus on Australia. Several aspects of temperature measurement were discussed to highlight the need for temperature reconstruction. Factors contributing to a discontinuous and non-homogeneous temperature record, including changes in measurement methods and the surface area surrounding weather stations over history, were discussed in Section 2.

The standard approach of dealing with unwanted artefacts present in the temperature record is homogenisation. Homogenisation methods and the official temperature reconstruction of the Australian BoM known as ACORN-SAT were presented with the associated historical temperature trends in Section 3.

The aim of the study presented in this report is to investigate alternative temperature reconstruction techniques, starting from the raw measurements made available by the BoM. Part of such an investigation is finding methods that can be used as benchmarks, that are easily reproducible for any geographical area. To that end, an overview of a number of existing spatial interpolation techniques were given in Section 4. These techniques are commonly used in the estimation of spatially-distributed environmental variables, including temperature. Inverse-distance methods were found to be suitable benchmarks for the development of alternative temperature reconstruction techniques.

As the temperature reconstruction methods considered in this report use raw measurements, quality control of the data must first be performed to remove potentially invalid measurements. Quality control techniques including control charts and nearest neighbours were presented in Section 5. Control charting is a statistical process control technique commonly used in manufacturing industries, which may also be applied to evaluate the validity of individual temperature series in isolation. The nearest-neighbour technique was presented as an alternative to identify outliers in time series by comparing individual values with their geographic neighbours for the same time index.

Two temperature reconstruction techniques alternative to the homogenisation process were presented in this report, including the nearest-neighbour technique in Section 6, and the neural network technique in Section 8.

The nearest-neighbour technique was used to infill all monthly $T_{\max}$ and $T_{\min}$ series across Australia and for each state separately, from which long-term trends were determined. The infilled series and trends were also compared with ACORN-SAT and found to be

similar. The similarity between ACORN-SAT and the nearest-neighbour anomaly trends was found to be particularly striking for Victoria, which has arguably the best station distribution in Australia. The areas between weather stations are small across Victoria, resulting in the raw data already being approximately area weighted. This is probably the reason why the area-weighted homogenised ACORN-SAT trend and the trend derived from the nearest-neighbour infilled raw data was found to be nearly identical for Victoria.

In preparation for the neural network reconstruction, background AI and neural network theory were given in Section 7, with further mathematical derivations in Appendix A. The reconstruction technique was then presented in Section 8, with the focus on an area around Deniliquin, a town in New South Wales. The technique was compared with inverse distance methods when several individual $T_{\max}$ series were estimated. Overall, the neural network technique was shown to perform better than IDW when temperature data at unsampled locations are to be estimated. Further development and performance evaluation are required before the neural network technique could be used to estimate historical temperature trends over larger regions.

10 Future work

Areas of future work to improve the temperature reconstruction methods presented in this report are discussed below.

- The ANN technique could be developed further to improve its performance. Other ANN structures and training strategies could be investigated. Performance could possibly be improved by training the network on entire time series at each location, instead of only one temperature sample. Other inputs could also be included to further guide the training process.

- Further experimentation is needed to improve station selection for infilling and estimation purposes. In this report, the closest N neighbouring stations were included to perform estimation. An improved method may be to rather select stations based on their correlation with neighbours, and excluding those stations that don't correlate well (even though they may be located closer).

- The estimation techniques considered in this report should also be tested on other and larger areas. Ultimately, a continental reconstruction should be performed.

Other aspects that need to be investigated further are considered below.

- The effect of UHIs (discussed in Section 2.2) on the temperature record should be considered further. The BoM excludes 8 urban sites (as discussed in Section 3.2) when calculating regional averages, whereas the alternative reconstructions presented in this report included all available data. It is also possible that other sites may be affected by UHI and a mechanism to detect and remove such artificial warming should be considered.
- The effect of changing from classical thermometers to AWSs (as discussed in Section 2.1) should be investigated further, as this may also contribute to non-climatic warming. Excluding AWS measurements is one approach, although this will remove a significant amount of recently-recorded data from the analysis. A separate study on the effect of the change should be conducted, which will involve statistical analysis of parallel data (classical and AWS measurements at the same location).
- The effect of station relocations should also be revisited. The station locations provided in the BoM weather station directory [108] were used to train the ANN in this report. These locations are presumably the latest station locations. In some cases stations were moved, although the same station ID was retained. Further details are obtainable in the basic climatological station metadata (e.g. [109]), which should be investigated individually for each weather station to uncover more detail regarding historical site locations.

The three above-mentioned aspects may have a significant impact on the historical temperature trends hidden in the raw record. It is possible that the alternative reconstructions presented in this report will provide trends similar to ACORN-SAT, as already indicated in this report. All these reconstructions start from the same raw data and similar trends are therefore not unexpected. Further analysis of the raw record may therefore be an important next step.

11 Acknowledgements

This research was funded by the B. Macfie Family Foundation.

A Mathematical derivations for online ANN training

This appendix presents mathematical derivations which were necessary during the development of software to train the networks presented in this report, using online ML algorithms. Only a 2-2-2 ANN with the classical log-sigmoid transfer function as shown in Fig. 60 is considered here. Both forward propagation and backpropagation are considered, after which a numerical example based on [110] is presented.

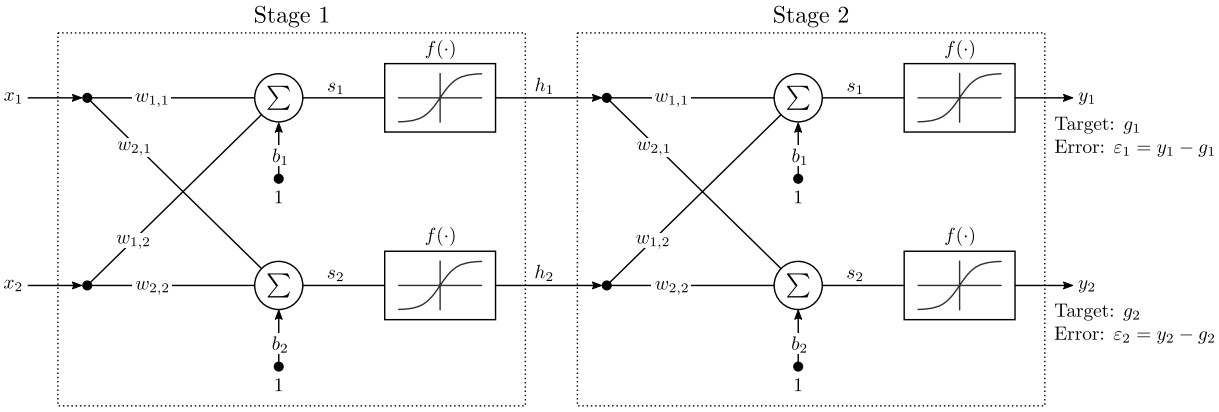


Figure 60: 2-2-2 ANN with initial weights, input and target values of one training example.

A.1 Forward propagation

The first part of training is to propagate the input example forward through the network to calculate the output values. Using matrix notation, the outputs of the summation units in stage 1 can be written as

$$\mathbf{s}^{(1)} = \mathbf{w}^{(1)}\mathbf{x} + \mathbf{b}^{(1)} \tag{107}$$

$$\therefore \begin{bmatrix} s_1^{(1)} \\ s_2^{(1)} \end{bmatrix} = \begin{bmatrix} w_{1,1}^{(1)} & w_{1,2}^{(1)} \\ w_{2,1}^{(1)} & w_{2,2}^{(1)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \end{bmatrix} \tag{108}$$

The output of stage 1 is then

$$\mathbf{h} = f\left(\mathbf{s}^{(1)}\right) \tag{109}$$

$$\therefore \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} \text{logsig}\left(s_1^{(1)}\right) \\ \text{logsig}\left(s_2^{(1)}\right) \end{bmatrix} \tag{110}$$

with $\text{logsig}(x) = {}^{1}/_{(1+\mathrm{e}^{-x})}$.

As the network shown in Fig. 60 consists of two identical stages, the stage 1 output $\mathbf{h}$ is

used as stage 2 input. The summation outputs in stage 2 can therefore be obtained as

$$\mathbf{s}^{(2)} = \mathbf{w}^{(2)}\mathbf{h} + \mathbf{b}^{(2)} \tag{111}$$

$$\therefore \begin{bmatrix} s_1^{(2)} \\ s_2^{(2)} \end{bmatrix} = \begin{bmatrix} w_{1,1}^{(2)} & w_{1,2}^{(2)} \\ w_{2,1}^{(2)} & w_{2,2}^{(2)} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \end{bmatrix} \tag{112}$$

The final network output is then

$$\mathbf{y} = f\left(\mathbf{s}^{(2)}\right) \tag{113}$$

$$\therefore \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \text{logsig}\left(s_1^{(2)}\right) \\ \text{logsig}\left(s_2^{(2)}\right) \end{bmatrix} \tag{114}$$

A.2 Backpropagation

Finding the network weights that will minimise the output error entails determining the influence of each weight on the output, as mentioned in Section 7.4.4. This influence is expressed in terms of the gradient, or collection of directional derivatives $\frac{\partial \varepsilon}{\partial w_z}$, which is a measure of change in the error ε relative to a change in any given weight w_z, assuming everything else in the network remains unchanged.

Suppose the i^{th} output error (or target difference) of the ANN is defined as

$$\varepsilon_i = y_i - g_i \tag{115}$$

with y_i and g_i respectively the calculated and target output, as indicated in Fig. 60. The total error for the 2-2-2 ANN can then be expressed as the MSE

$$\varepsilon_{\text{t}} = \frac{1}{2}\varepsilon_1^2 + \frac{1}{2}\varepsilon_2^2 \tag{116}$$

$$= \frac{1}{2}\left(y_1 - g_1\right)^2 + \frac{1}{2}\left(y_2 - g_2\right)^2 \tag{117}$$

which is the loss function that will be used here.[26]

The partial derivative $\frac{\partial \varepsilon_{\text{t}}}{\partial w_z}$ can be calculated for every w_z in the network by propagating the error backwards through the network, by virtue of the chain rule of Calculus [98]. This concept will subsequently be illustrated through mathematical derivations and then summarised in matrix form for each stage of the network shown in Fig. 60.

[26] The loss function is often written as half an error value squared (irrespective the number of network outputs) to simplify the maths used to derive the error gradients, as the derivative of the loss function is involved. For the example considered here, the error ε_{t} is already in a convenient form, an is equivalent to the MSE.

A.2.1 Stage 2 derivation

Starting from output y_1, the error derivative w.r.t. weight w_z (either $w_{1,1}$, $w_{1,2}$ or b_1) in stage 2 of the network shown in Fig. 60 can be calculated using the above equations as follows.

$$\frac{\partial \varepsilon_\mathrm{t}}{\partial w_{z,y_1}} = \frac{\partial}{\partial w_z}\left\{\frac{1}{2}\left(y_1 - g_1\right)^2 + \frac{1}{2}\left(y_2 - g_2\right)^2\right\} \tag{118}$$

$$= \frac{\partial}{\partial w_z}\left\{\frac{1}{2}\left(y_1 - g_1\right)^2\right\} + \cancelto{0}{\frac{\partial}{\partial w_z}\left\{\frac{1}{2}\left(y_2 - g_2\right)^2\right\}} \tag{119}$$

with the second term zero as y_2 is not a function of w_z considered here, and g_2 is constant. The output y_1 is a function s_1, which in turn is a function of w_z, as indicated in Fig. 60. Using the chain rule, (119) can therefore be developed further as

$$\frac{\partial \varepsilon_\mathrm{t}}{\partial w_{z,y_1}} = 2 \times \frac{1}{2}(y_1 - g_1)\frac{\partial y_1}{\partial w_z} \tag{120}$$

$$= \varepsilon_1 \cdot \frac{\partial y_1}{\partial s_1} \cdot \frac{\partial s_1}{\partial w_z} \tag{121}$$

Since $y_1 = f(s_1) = \mathrm{logsig}(s_1)$, the derivative of the log-sigmoid function needs to be determined, which can be achieved using the quotient rule of Calculus and partial-fraction expansion as

$$f'(s_1) = \frac{\partial}{\partial s_1}\mathrm{logsig}(s_1) \tag{122}$$

$$= \frac{\partial}{\partial s_1}\left\{\frac{1}{1+\mathrm{e}^{-s_1}}\right\} \tag{123}$$

$$= \frac{\mathrm{e}^{-s_1}}{\left(1+\mathrm{e}^{-s_1}\right)^2} \tag{124}$$

$$= \frac{1}{1+\mathrm{e}^{-s_1}} - \frac{1}{\left(1+\mathrm{e}^{-s_1}\right)^2} \tag{125}$$

$$= \frac{1}{1+\mathrm{e}^{-s_1}}\left(1 - \frac{1}{1+\mathrm{e}^{-s_1}}\right) \tag{126}$$

$$= y_1\left(1 - y_1\right) \tag{127}$$

by noting that $y_1 = f_1(s_1) = \frac{1}{1+\mathrm{e}^{-s_1}}$. By plugging (127) back into (121), the error derivatives can then be expressed as

$$\frac{\partial \varepsilon_\mathrm{t}}{\partial w_{z,y_1}} = \varepsilon_1 y_1\left(1 - y_1\right)\frac{\partial}{\partial w_z}\left\{w_{1,1}h_1 + w_{1,2}h_2 + b_1\right\} \tag{128}$$

$$= \delta_1 \frac{\partial}{\partial w_z}\left\{w_{1,1}h_1 + w_{1,2}h_2 + b_1\right\} \tag{129}$$

with the node delta expression

$$\delta_1 = \varepsilon_1 f'(s_1) = \varepsilon_1 y_1(1 - y_1) \tag{130}$$

The error derivatives for each weight associated with y_1 in stage 2 can then be written as

$$\begin{aligned} \frac{\partial \varepsilon_t}{\partial w_{1,1}} &= \delta_1 h_1 \quad \text{if} \quad w_z = w_{1,1} \\ \frac{\partial \varepsilon_t}{\partial w_{1,2}} &= \delta_1 h_2 \quad \text{if} \quad w_z = w_{1,2} \\ \frac{\partial \varepsilon_t}{\partial b_1} &= \delta_1 \quad \text{if} \quad w_z = b_1 \end{aligned} \tag{131}$$

as illustrated in Fig. 61 for each weight.

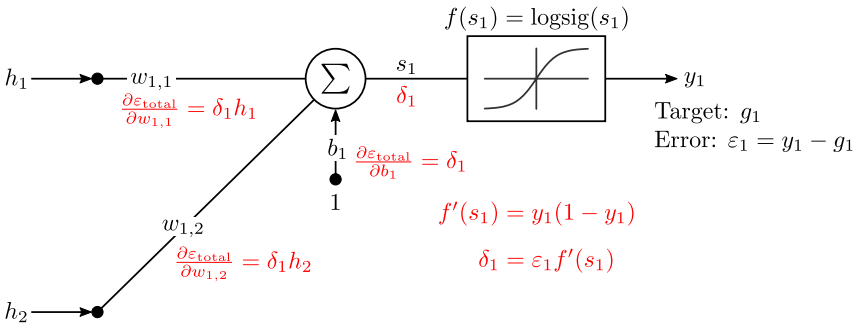


Figure 61: Backpropagation calculations (in red) for stage 2.

The same approach can be followed to calculate the derivatives of the other weights, associated with output y_2, in stage 2:

$$\frac{\partial \varepsilon_t}{\partial w_{z,y_2}} = \frac{\partial \varepsilon_t}{\partial y_2} \cdot \frac{\partial y_2}{\partial s_2} \cdot \frac{\partial s_2}{\partial w_z} \tag{132}$$

$$= \frac{\partial}{\partial y_2} \left\{ \frac{1}{2}\varepsilon_1^2 + \frac{1}{2}\varepsilon_2^2 \right\} \cdot \frac{\partial}{\partial s_2} \text{logsig}(s_2) \cdot \frac{\partial}{\partial w_z} \{w_{2,1}h_1 + w_{2,2}h_2 + b_2\} \tag{133}$$

$$= \varepsilon_2 y_2(1 - y_2) \frac{\partial}{\partial w_z} \{w_{2,1}h_1 + w_{2,2}h_2 + b_2\} \tag{134}$$

$$= \delta_2 \frac{\partial}{\partial w_z} \{w_{2,1}h_1 + w_{2,2}h_2 + b_2\} \tag{135}$$

with

$$\delta_2 = \varepsilon_2 f'(s_2) = \varepsilon_2 y_2(1 - y_2) \tag{136}$$

The derivative w.r.t. each weight associated with y_2 can then be expressed as

$$\begin{aligned} \frac{\partial \varepsilon_t}{\partial w_{2,1}} &= \delta_2 h_1 \quad \text{if} \quad w_z = w_{2,1} \\ \frac{\partial \varepsilon_t}{\partial w_{2,2}} &= \delta_2 h_2 \quad \text{if} \quad w_z = w_{2,2} \\ \frac{\partial \varepsilon_t}{\partial b_2} &= \delta_2 \quad \text{if} \quad w_z = b_2 \end{aligned} \tag{137}$$

A.2.2 Stage 2 matrix summary

If the weights in stage 2 of the ANN shown in Fig. 60 are written as

$$\mathbf{w} = \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{138}$$

the gradients can be expressed as

$$\nabla\varepsilon(\mathbf{w}) = \begin{bmatrix} \dfrac{\partial\varepsilon_t}{\partial w_{1,1}} & \dfrac{\partial\varepsilon_t}{\partial w_{1,2}} \\ \dfrac{\partial\varepsilon_t}{\partial w_{2,1}} & \dfrac{\partial\varepsilon_t}{\partial w_{2,2}} \end{bmatrix} \text{ and } \nabla\varepsilon(\mathbf{b}) = \begin{bmatrix} \dfrac{\partial\varepsilon_t}{\partial b_1} \\ \dfrac{\partial\varepsilon_t}{\partial b_2} \end{bmatrix} \tag{139}$$

which can be written compactly from (131) and (137) as

$$\nabla\varepsilon(\mathbf{w}) = \begin{bmatrix} \delta_1 h_1 & \delta_1 h_2 \\ \delta_2 h_1 & \delta_2 h_2 \end{bmatrix} \text{ and } \nabla\varepsilon(\mathbf{b}) = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \bar{\delta} \tag{140}$$

which can further be factorised (in terms of matrix multiplication) as

$$\nabla\varepsilon(\mathbf{w}) = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \begin{bmatrix} h_1 & h_2 \end{bmatrix} \tag{141}$$

$$= \bar{\delta}\mathbf{h}^T \tag{142}$$

The node deltas defined in (130) and (136) can be written in matrix form as

$$\bar{\delta} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \circ \begin{bmatrix} f'\left(s_1^{(2)}\right) \\ f'\left(s_2^{(2)}\right) \end{bmatrix} \tag{143}$$

or compactly as

$$\bar{\delta} = \bar{\varepsilon} \circ \mathbf{f}'^{(2)} \tag{144}$$

with $\bar{\varepsilon}$ the error column vector, $\circ$ the Hadamard (element-wise matrix multiplication) operator, and the stage-2 transfer function derivative column vector

$$\mathbf{f}'^{(2)} = \begin{bmatrix} f'\left(s_1^{(2)}\right) \\ f'\left(s_2^{(2)}\right) \end{bmatrix} = \begin{bmatrix} y_1(1-y_1) \\ y_2(1-y_2) \end{bmatrix} \tag{145}$$

The weights can then be updated after completion of stage 1 processing using

$$\mathbf{w}^+ = \mathbf{w} - \eta\nabla\varepsilon(\mathbf{w}) \text{ and } \mathbf{b}^+ = \mathbf{b} - \eta\bar{\delta} \tag{146}$$

A.2.3 Stage 1 derivation

Calculating the gradient components for stage 1 associated with output h_1 takes the same form as (132):

$$\frac{\partial \varepsilon_{\mathrm{t}}}{\partial w_{z,h_1}} = \frac{\partial \varepsilon_{\mathrm{t}}}{\partial h_1} \cdot \frac{\partial h_1}{\partial s_1} \cdot \frac{\partial s_1}{\partial w_z} \tag{147}$$

with $\frac{\partial \varepsilon_{\mathrm{t}}}{\partial h_1}$ expanded using (117) as

$$\frac{\partial \varepsilon_{\mathrm{t}}}{\partial h_1} = \frac{\partial}{\partial h_1} \left\{ \frac{1}{2} \left(y_1 - g_1 \right)^2 + \frac{1}{2} \left(y_2 - g_2 \right)^2 \right\} \tag{148}$$

$$= (y_1 - g_1) \frac{\partial y_1}{\partial h_1} + (y_2 - g_2) \frac{\partial y_2}{\partial h_1} \tag{149}$$

$$= \varepsilon_1 \frac{\partial y_1}{\partial s_1^{(2)}} \frac{\partial s_1^{(2)}}{\partial h_1} + \varepsilon_2 \frac{\partial y_2}{\partial s_2^{(2)}} \frac{\partial s_2^{(2)}}{\partial h_1} \tag{150}$$

$$= \varepsilon_1 f' \left(s_1^{(2)} \right) w_{1,1}^{(2)} + \varepsilon_2 f' \left(s_2^{(2)} \right) w_{2,1}^{(2)} \tag{151}$$

$$= \delta_1 w_{1,1}^{(2)} + \delta_2 w_{2,1}^{(2)} \tag{152}$$

written by inspection from stage 2 in Fig. 60, with the superscript $\cdot^{(2)}$ signifying stage 2, and substituting the node deltas defined in (130) and (136).

By noting further that

$$h_1 = f(s_1) = \mathrm{logsig}(s_1) \tag{153}$$

$$s_1 = w_{1,1} x_1 + w_{1,2} x_2 + b_1 \tag{154}$$

the error derivative of (147) can be expressed as

$$\frac{\partial \varepsilon_{\mathrm{t}}}{\partial w_{z,h_1}} = \left(\delta_1 w_{1,1}^{(2)} + \delta_2 w_{2,1}^{(2)} \right) \cdot f'(s_1) \cdot \frac{\partial}{\partial w_z} \left\{ w_{1,1} x_1 + w_{1,2} x_2 + b_1 \right\} \tag{155}$$

$$= \delta_{\mathrm{h},1} \frac{\partial}{\partial w_z} \left\{ w_{1,1} x_1 + w_{1,2} x_2 + b_1 \right\} \tag{156}$$

with the hidden-layer node delta defined as

$$\delta_{\mathrm{h},1} = \left(\delta_1 w_{1,1}^{(2)} + \delta_2 w_{2,1}^{(2)} \right) f'(s_1) = \left(\delta_1 w_{1,1}^{(2)} + \delta_2 w_{2,1}^{(2)} \right) h_1 (1 - h_1) \tag{157}$$

Following the same approach, the gradient components for stage 1 associated with output h_2 can be expressed in the exact same form as (156):

$$\frac{\partial \varepsilon_{\mathrm{t}}}{\partial w_{z,h_2}} = \left(\delta_1 w_{1,2}^{(2)} + \delta_2 w_{2,2}^{(2)} \right) \cdot f'(s_2) \cdot \frac{\partial}{\partial w_z} \left\{ w_{2,1} x_1 + w_{2,2} x_2 + b_2 \right\} \tag{158}$$

$$= \delta_{\mathrm{h},2} \frac{\partial}{\partial w_z} \left\{ w_{2,1} x_1 + w_{2,2} x_2 + b_2 \right\} \tag{159}$$

with the hidden-layer node delta defined as

$$\delta_{\mathrm{h},2} = \left(\delta_1 w_{1,2}^{(2)} + \delta_2 w_{2,2}^{(2)} \right) f'(s_2) = \left(\delta_1 w_{1,2}^{(2)} + \delta_2 w_{2,2}^{(2)} \right) h_2 (1 - h_2) \tag{160}$$

A.2.4 Stage 1 matrix summary

The stage 1 gradients can be written in the same form as (139), using (156) and (159) as

$$\nabla\varepsilon(\mathbf{w}) = \begin{bmatrix} \delta_{\mathrm{h},1}x_1 & \delta_{\mathrm{h},1}x_2 \\ \delta_{\mathrm{h},2}x_1 & \delta_{\mathrm{h},2}x_2 \end{bmatrix} \text{ and } \nabla\varepsilon(\mathbf{b}) = \begin{bmatrix} \delta_{\mathrm{h},1} \\ \delta_{\mathrm{h},2} \end{bmatrix} = \bar{\delta}_{\mathrm{h}} \tag{161}$$

which can be factorised (in terms of matrix multiplication) similarly as in Section A.2.2 as

$$\nabla\varepsilon(\mathbf{w}) = \begin{bmatrix} \delta_{\mathrm{h},1} \\ \delta_{\mathrm{h},2} \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix} \tag{162}$$

$$= \bar{\delta}_{\mathrm{h}}\mathbf{x}^T \tag{163}$$

The hidden-layer node deltas defined in (157) and (160) can also be expressed conveniently in matrix form as

$$\bar{\delta}_{\mathrm{h}} = \begin{bmatrix} \delta_{\mathrm{h},1} \\ \delta_{\mathrm{h},2} \end{bmatrix} = \begin{bmatrix} w_{1,1}^{(2)} & w_{2,1}^{(2)} \\ w_{1,2}^{(2)} & w_{2,2}^{(2)} \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \circ \begin{bmatrix} f'\left(s_1^{(1)}\right) \\ f'\left(s_2^{(1)}\right) \end{bmatrix} \tag{164}$$

or compactly as

$$\bar{\delta}_{\mathrm{h}} = \left[\mathbf{w}^{(2)}\right]^T \bar{\delta} \circ \mathbf{f}'^{(1)} \tag{165}$$

with $\mathbf{w}^{(2)}$ the weight matrix of stage 2, $\bar{\delta}$ the node delta column vector of stage 2, $\circ$ the Hadamard (element-wise matrix multiplication) operator, and the stage-1 transfer function derivative column vector

$$\mathbf{f}'^{(1)} = \begin{bmatrix} f'\left(s_1^{(1)}\right) \\ f'\left(s_2^{(1)}\right) \end{bmatrix} = \begin{bmatrix} h_1(1-h_1) \\ h_2(1-h_2) \end{bmatrix} \tag{166}$$

Finally, the stage-1 weights can be updated using

$$\mathbf{w}^+ = \mathbf{w} - \eta\nabla\varepsilon(\mathbf{w}) \text{ and } \mathbf{b}^+ = \mathbf{b} - \eta\bar{\delta}_{\mathrm{h}} \tag{167}$$

A.3 Numerical example

A numerical example using the equations derived in this Appendix is subsequently presented. The ANN structure shown in Fig. 60 is used in this example, with the values of trainable parameters and training data from [110].

A single training data example is presented to the network. The input data is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.05 \\ 0.1 \end{bmatrix} \tag{168}$$

with associated target data

$$\mathbf{g} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} 0.01 \\ 0.99 \end{bmatrix} \tag{169}$$

The initial weight values of stage 1 are

$$\mathbf{w}^{(1)} = \begin{bmatrix} w_{1,1}^{(1)} & w_{1,2}^{(1)} \\ w_{2,1}^{(1)} & w_{2,2}^{(1)} \end{bmatrix} = \begin{bmatrix} 0.15 & 0.2 \\ 0.25 & 0.3 \end{bmatrix} \text{ and } \mathbf{b}^{(1)} = \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \end{bmatrix} = \begin{bmatrix} 0.35 \\ 0.35 \end{bmatrix} \tag{170}$$

and of stage 2:

$$\mathbf{w}^{(2)} = \begin{bmatrix} w_{1,1}^{(2)} & w_{1,2}^{(2)} \\ w_{2,1}^{(2)} & w_{2,2}^{(2)} \end{bmatrix} = \begin{bmatrix} 0.4 & 0.45 \\ 0.5 & 0.55 \end{bmatrix} \text{ and } \mathbf{b}^{(2)} = \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.6 \end{bmatrix} \tag{171}$$

A.3.1 Forward propagation

Using the equations presented in Section A.1, the stage 1 sum outputs are

$$\mathbf{s}^{(1)} = \mathbf{w}^{(1)}\mathbf{x} + \mathbf{b}^{(1)} \tag{172}$$

$$\therefore \begin{bmatrix} s_1^{(1)} \\ s_2^{(1)} \end{bmatrix} = \begin{bmatrix} 0.3775 \\ 0.3925 \end{bmatrix} \tag{173}$$

and the stage 1 outputs are

$$\mathbf{h} = f\left(\mathbf{s}^{(1)}\right) \tag{174}$$

$$\therefore \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} 0.593269992107187 \\ 0.596884378259767 \end{bmatrix} \tag{175}$$

The stage 2 sum unit outputs are

$$\mathbf{s}^{(2)} = \mathbf{w}^{(2)}\mathbf{h} + \mathbf{b}^{(2)} \tag{176}$$

$$\therefore \begin{bmatrix} s_1^{(2)} \\ s_2^{(2)} \end{bmatrix} = \begin{bmatrix} 1.105905967059770 \\ 1.224921404096465 \end{bmatrix} \tag{177}$$

and the final outputs are

$$\mathbf{y} = f\left(\mathbf{s}^{(2)}\right) \tag{178}$$

$$\therefore \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0.751365069552316 \\ 0.772928465321463 \end{bmatrix} \tag{179}$$

A.3.2 Backpropagation

The first step is calculating the error, which can be expressed as

$$\bar{\varepsilon} = \mathbf{y} - \mathbf{g} \tag{180}$$

$$= \begin{bmatrix} 0.751365069552316 \\ 0.772928465321463 \end{bmatrix} - \begin{bmatrix} 0.01 \\ 0.99 \end{bmatrix} = \begin{bmatrix} 0.741365069552316 \\ -0.217071534678537 \end{bmatrix} \tag{181}$$

Stage 2

Using the equations presented in Section A.2.2, the stage-2 transfer function derivative can be calculated as

$$\mathbf{f}'^{(2)} = \begin{bmatrix} y_1(1-y_1) \\ y_2(1-y_2) \end{bmatrix} = \begin{bmatrix} 0.186815601808959 \\ 0.175510052817271 \end{bmatrix} \tag{182}$$

and the stage-2 node deltas as

$$\bar{\delta} = \bar{\varepsilon} \circ \mathbf{f}'^{(2)} = \begin{bmatrix} 0.138498561628557 \\ -0.038098236516556 \end{bmatrix} \tag{183}$$

The stage-2 gradients can then be expressed as

$$\nabla\varepsilon(\mathbf{w}) = \bar{\delta}\mathbf{h}^T = \begin{bmatrix} 0.082167040564231 & 0.082667627847533 \\ -0.022602540477475 & -0.022740242215978 \end{bmatrix} \tag{184}$$

and

$$\nabla\varepsilon(\mathbf{b}) = \bar{\delta} \tag{185}$$

Stage 1

Using the equations presented in Section A.2.4, the stage-1 transfer function derivative can be calculated as

$$\mathbf{f}'^{(1)} = \begin{bmatrix} h_1(1-h_1) \\ h_2(1-h_2) \end{bmatrix} = \begin{bmatrix} 0.241300708572325 \\ 0.240613417249218 \end{bmatrix} \tag{186}$$

and the stage-1 node deltas as

$$\bar{\delta}_{\mathrm{h}} = \left[\mathbf{w}^{(2)}\right]^T \bar{\delta} \circ \mathbf{f}'^{(1)} = \begin{bmatrix} 0.008771354689487 \\ 0.009954254705217 \end{bmatrix} \tag{187}$$

The stage-1 gradients can then be expressed as

$$\nabla\varepsilon\left(\mathbf{w}^{(1)}\right) = \bar{\delta}_{\mathrm{h}}\mathbf{x}^T = \begin{bmatrix} 0.000438567734474 & 0.000877135468949 \\ 0.000497712735261 & 0.000995425470522 \end{bmatrix} \tag{188}$$

and

$$\nabla\varepsilon\left(\mathbf{b}^{(1)}\right) = \bar{\delta}_{\mathrm{h}} \tag{189}$$

Updating the weights

Using a step size or learning rate of $\eta = 0.5$, the stage-1 weights can be updated as

$$\mathbf{w}^{(1)+} = \mathbf{w}^{(1)} - \eta\nabla\varepsilon\left(\mathbf{w}^{(1)}\right) \tag{190}$$

$$= \begin{bmatrix} 0.15 & 0.2 \\ 0.25 & 0.3 \end{bmatrix} - 0.5 \begin{bmatrix} 0.000438567734474 & 0.000877135468949 \\ 0.000497712735261 & 0.000995425470522 \end{bmatrix} \tag{191}$$

$$= \begin{bmatrix} 0.149780716132763 & 0.199561432265526 \\ 0.249751143632370 & 0.299502287264739 \end{bmatrix} \tag{192}$$

and

$$\mathbf{b}^{(1)+} = \mathbf{b}^{(1)} - \eta\bar{\delta}_{\mathrm{h}} \tag{193}$$

$$= \begin{bmatrix} 0.35 \\ 0.35 \end{bmatrix} - 0.5 \begin{bmatrix} 0.008771354689487 \\ 0.009954254705217 \end{bmatrix} \tag{194}$$

$$= \begin{bmatrix} 0.345614322655256 \\ 0.345022872647391 \end{bmatrix} \tag{195}$$

The stage-2 weights can similarly be updated as

$$\mathbf{w}^{(2)+} = \mathbf{w}^{(2)} - \eta\nabla\varepsilon\left(\mathbf{w}^{(2)}\right) \tag{196}$$

$$= \begin{bmatrix} 0.4 & 0.45 \\ 0.5 & 0.55 \end{bmatrix} - 0.5 \begin{bmatrix} 0.082167040564231 & 0.082667627847533 \\ -0.022602540477475 & -0.022740242215978 \end{bmatrix} \tag{197}$$

$$= \begin{bmatrix} 0.358916479717885 & 0.408666186076233 \\ 0.511301270238738 & 0.561370121107989 \end{bmatrix} \tag{198}$$

and

$$\mathbf{b}^{(2)+} = \mathbf{b}^{(2)} - \eta\bar{\delta} \tag{199}$$

$$= \begin{bmatrix} 0.6 \\ 0.6 \end{bmatrix} - 0.5 \begin{bmatrix} 0.138498561628557 \\ -0.038098236516556 \end{bmatrix} \tag{200}$$

$$= \begin{bmatrix} 0.530750719185721 \\ 0.619049118258278 \end{bmatrix} \tag{201}$$

Fig. 62 shows how the errors are reduced over 100 training iterations, using the single training data example with input defined in (168) and target output defined in (169).

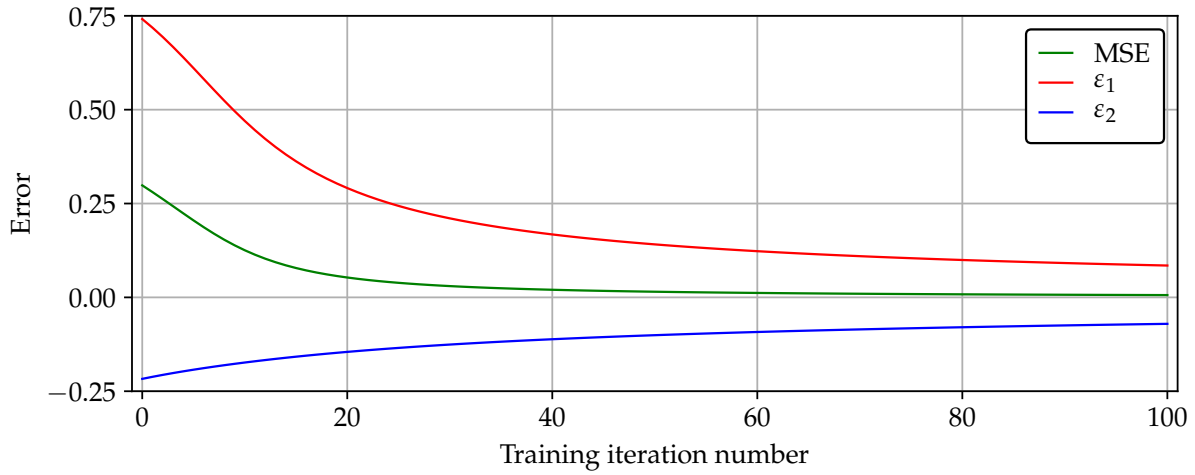


Figure 62: Error behaviour during training when using the single training data example.

References

[1] D. Sherry, "Thermoscopes, thermometers, and the foundations of measurement," *Studies in History and Philosophy of Science Part A*, vol. 42, no. 4, pp. 509–524, 2011.

[2] G. Manley, "Central England temperatures: monthly means 1659 to 1973," *Quarterly Journal of the Royal Meteorological Society*, vol. 100, no. 425, pp. 389–405, 1974.

[3] B. Trewin, "Techniques involved in developing the Australian Climate Observations Reference Network - Surface Air Temperature (ACORN-SAT) dataset," The Centre for Australian Weather and Climate Research: A partnership between CSIRO and the Bureau of Meteorology, Tech. Rep. CAWCR Technical Report No. 049, March 2012, Accessed 12 December 2018. [Online]. Available: http://cawcr.gov.au/technical-reports/CTR_049.pdf

[4] J. Hansen, R. Ruedy, M. Sato, and K. Lo, "Global surface temperature change," *Reviews of Geophysics*, vol. 48, no. 4, 2010.

[5] C. P. Morice, J. J. Kennedy, N. A. Rayner, and P. D. Jones, "Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set," *Journal of Geophysical Research: Atmospheres*, vol. 117, no. D8, 2012.

[6] National Oceanic and Atmospheric Administration: National Centers for Environmental Information, "Climate at a Glance: Global Time Series," 2019, Accessed: 15 February 2019. [Online]. Available: https://www.ncdc.noaa.gov/cag/global/time-series/globe/land_ocean/ytd/12/1880-2016

[7] R. Rohde, R. Muller, R. Jacobsen, E. Muller *et al.*, "A new estimate of the average Earth surface land temperature spanning 1753 to 2011," *Geoinformatics and Geostatistics: An Overview*, vol. 1, pp. 1–7, 2013.

[8] National Aeronautics and Space Administration: Goddard Institute for Space Studies, "GISS Surface Temperature Analysis: Global Annual Mean Surface Air Temperature Change," Accessed: 14 February 2019. [Online]. Available: https://data.giss.nasa.gov/gistemp/graphs/

[9] F. C. Ljungqvist, "Temperature proxy records covering the last two millennia: a tabular and visual overview," *Geografiska Annaler: Series A, Physical Geography*, vol. 91, no. 1, pp. 11–29, 2009.

[10] J. T. Houghton, G. J. Jenkins, and J. Ephraums, *Climate Change: The IPCC Scientific Assessment*. Cambridge University Press, 1990.

[11] S. Desprat, M. Goñi, and M. Loutre, "Revealing climatic variability of the last three millennia in northwestern Iberia using pollen influx data," *Earth and Planetary Science Letters*, vol. 213, no. 1-2, pp. 63–78, 2003.

[12] T. J. Crowley and T. S. Lowery, "How warm was the medieval warm period?" *AMBIO: A Journal of the Human Environment*, vol. 29, no. 1, pp. 51–54, 2000.

[13] W. S. Broecker, "Was the medieval warm period global?" *Science*, vol. 291, no. 5508, pp. 1497–1499, 2001.

[14] D. J. MacKay, *Sustainable Energy - without the hot air*. UIT Cambridge, 2008, Available free online at www.withouthotair.com.

[15] S. M. Kay, *Fundamentals of statistical signal processing: Practical algorithm development*. Pearson Education, 2013, vol. 3.

[16] J. D. Shakun, P. U. Clark, F. He, S. A. Marcott *et al.*, "Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation," *Nature*, vol. 484, no. 7392, p. 49, 2012.

[17] J. Abbot and J. Nicol, "The contribution of carbon dioxide to global warming," in *Climate Change: The Facts 2017*, J. J. Marohasy, Ed. Melbourne, Australia: Institute of Public Affairs, 2017, ch. 11, pp. 282–296.

[18] S. Arrhenius, "On the influence of carbonic acid in the air upon the temperature of the ground," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 41, no. 251, pp. 237–276, 1896.

[19] J. W. Abbot and J. J. Marohasy, "The application of machine learning for evaluating anthropogenic versus natural climate change," *GeoResJ*, vol. 14, pp. 36–46, 2017.

[20] N. Scafetta, "Understanding climate change in terms of natural variability," in *Climate Change: The Facts 2017*, J. J. Marohasy, Ed. Melbourne, Australia: Institute of Public Affairs, 2017, ch. 3, pp. 39–58.

[21] W. Soon and S. Baliunas, "A brief review of the sun-climate connection, with a new insight concerning water vapour," in *Climate Change: The Facts 2017*, J. J. Marohasy, Ed. Melbourne, Australia: Institute of Public Affairs, 2017, ch. 11, pp. 163–177.

[22] Scripps Institution of Oceanography: CO_2 Program, "Atmospheric CO_2 Data: Ice-Core Merged Products," 2019, Accessed: 18 February 2019. [Online]. Available: http://scrippsco2.ucsd.edu/data/atmospheric_co2/icecore_merged_products

[23] C. D. Keeling, S. C.Piper, R. B. Bacastow, M. Wahlen *et al.*, "Atmospheric CO_2 and $^{13}CO_2$ exchange with the terrestrial biosphere and oceans from 1978 to 2000: Observations and carbon cycle implications," in *A history of atmospheric CO_2 and its effects on plants, animals, and ecosystems*. Springer, 2005, pp. 83–113.

[24] C. MacFarling Meure, D. Etheridge, C. Trudinger, P. Steele *et al.*, "Law Dome CO_2, CH_4 and N_2O ice core records extended to 2000 years BP," *Geophysical Research Letters*, vol. 33, no. 14, 2006.

[25] IPCC Technical Support Unit (Working Group I), "Global warming of 1.5 °C," Intergovernmental panel on climate change, Tech. Rep. Special Report: Summary for Policymakers, January 2019, Accessed 18 February 2019. [Online]. Available: https://www.ipcc.ch/sr15/

[26] V. V. Zharkova, S. J. Shepherd, E. Popova, and S. I. Zharkov, "Heartbeat of the sun from principal component analysis and prediction of solar activity on a millenium timescale," *Scientific reports*, vol. 5, p. 15689, 2015.

[27] Judith Lean, NOAA/NGDC Paleoclimatology Program, Boulder CO, USA., "Solar Irradiance Reconstruction [1610-2000]," 2004, Accessed: 18 February 2019, IGBP PAGES/World Data Center for Paleoclimatology, Data Contribution Series # 2004-035. [Online]. Available: https://www1.ncdc.noaa.gov/pub/data/paleo/climate_forcing/solar_variability/lean2000_irradiance.txt

[28] Royal Observatory of Belgium: Sunspot Index and Long-term Solar Observations, "Total sunspot number (yearly mean total) [1700-2018]," 2019, Accessed: 18 February 2019, SILSO data is under CC BY-NC4.0 license (https://goo.gl/PXrLYd). [Online]. Available: http://www.sidc.be/silso/DATA/SN_y_tot_V2.0.txt

[29] J. A. Eddy, "The Maunder Minimum," *Science*, vol. 192, no. 4245, pp. 1189–1202, 1976, Table 1 data available at https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-indices/sunspot-numbers/ancient/eddy/documentation/eddy(1976)_science.txt.

[30] N. Gordon, A. Jonko, P. Forster, and K. Shell, "An observationally based constraint on the water-vapor feedback," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 22, pp. 12–435, 2013.

[31] M. Etminan, G. Myhre, E. Highwood, and K. Shine, "Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing," *Geophysical Research Letters*, vol. 43, no. 24, 2016.

[32] O. Sorokhtin, G. Chilingar, N. Sorokhtin, M. Liu *et al.*, "Jupiters effect on earths climate," *Environmental earth sciences*, vol. 73, no. 8, pp. 4091–4097, 2015.

[33] K. Ring, *The Lunar Code*. Random House New Zealand, 2006.

[34] U.S. Global Change Research Program, "Climate Science Special Report," 2017, Accessed: 19 February 2019. [Online]. Available: https://data.globalchange.gov/report/climate-science-special-report

[35] Australian Government: Bureau of Meteorology, "Long-term temperature record: ACORN-SAT," 2018, Accessed: 11 December 2018. [Online]. Available: http://www.bom.gov.au/climate/change/acorn-sat/documents/About_ACORN-SAT.pdf

[36] Australian Government: Bureau of Meteorology, "Australian stations measuring mean maximum air temperature," 2019, Accessed: 29 January 2019. [Online]. Available: http://www.bom.gov.au/climate/data/lists_by_element/alphaAUS_36.txt

[37] Australian Government: Bureau of Meteorology, "Ask BOM: How is temperature measured?" April 2016, Accessed: 6 December 2018. [Online]. Available: http://media.bom.gov.au/social/blog/916/ask-bom-how-is-temperature-measured/

[38] Australian Government: Bureau of Meteorology, "Observation of air temperature," June 2018, Accessed: 6 December 2018. [Online]. Available: http://www.bom.gov.au/climate/cdo/about/airtemp-measure.shtml

[39] *British Standard Specification for Industrial platinum resistance thermometer sensors*, British Standards Institution, June 1992, BS1904:1984.

[40] Bureau of Meteorology, "Australian Climate Observations Reference Network - Surface Air Temperature (ACORN-SAT): Observation practices," Australian Government, Tech. Rep., 2012, Accessed 6 December 2018. [Online]. Available: http://www.bom.gov.au/climate/change/acorn-sat/documents/ACORN-SAT_Observation_practices_WEB.pdf

[41] J. J. Marohasy, "Letter to the Chief Scientist," 2018, Accessed: 4 December 2018. [Online]. Available: https://jennifermarohasy.com/wp-content/uploads/2014/05/Marohasy-to-Finkel-20180504.pdf

[42] Australian Government: Bureau of Meteorology, "About Weather Station Directory," 2018, Accessed: 3 December 2018. [Online]. Available: http://www.bom.gov.au/climate/cdo/about/about-directory.shtml

[43] Australian Government: Bureau of Meteorology, "Date of AWS installation for Australian weather stations," 2018, Accessed: 3 December 2018. [Online]. Available: ftp://ftp.bom.gov.au/anon2/home/ncc/metadata/sitelists/aws_firstdate.xls

[44] Australian Government: Bureau of Meteorology, "Basic Climatological Station Metadata: Mildura Airport," 2018, Accessed: 4 December 2018. [Online]. Available: http://www.bom.gov.au/clim_data/cdio/metadata/pdf/siteinfo/IDCJMD0040.076031.SiteInfo.pdf

[45] J. J. Marohasy, "Temperature change at Rutherglen in South-East Australia," *New Climate*, 2016, http://dx.doi.org/10.22221/nc.2016.001. [Online]. Available: http://climatelab.com.au/wp-content/uploads/NW2016.001.PP_.Marohasy.pdf

[46] R. R. McKitrick and P. J. Michaels, "Quantifying the influence of anthropogenic surface processes and inhomogeneities on gridded global climate data," *Journal of Geophysical Research*, vol. 112, no. D24, 2007.

[47] Y. He, G. Jia, Y. Hu, and Z. Zhou, "Detecting urban warming signals in climate records," *Advances in Atmospheric Sciences*, vol. 30, no. 4, pp. 1143–1153, 2013.

[48] B. Trewin, "Exposure, instrumentation, and observing practice effects on land temperature measurements," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 1, no. 4, pp. 490–506, 2010.

[49] S. T. Buisan, C. Azorin-Molina, and Y. Jimenez, "Impact of two different sized stevenson screens on air temperature measurements," *International Journal of Climatology*, vol. 35, no. 14, pp. 4408–4416, 2015.

[50] Australian Government: Bureau of Meteorology, "Long-term temperature record: ACORN-SAT," 2018, Accessed: 10 December 2018. [Online]. Available: http://www.bom.gov.au/climate/change/acorn-sat/

[51] B. Trewin, "Effects of changes in algorithms used for the calculation of Australian mean temperature," *Australian Meteorological Magazine*, vol. 53, no. 1, 2004.

[52] B. Trewin, "A daily homogenized temperature data set for Australia," *International Journal of Climatology*, vol. 33, no. 6, pp. 1510–1529, 2013.

[53] S. Ribeiro, J. Caineta, and A. Costa, "Review and discussion of homogenisation methods for climate data," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 94, pp. 167–179, 2016.

[54] M. Basseville, I. V. Nikiforov *et al.*, *Detection of abrupt changes: theory and application*. Prentice Hall Englewood Cliffs, 1993, vol. 104. [Online]. Available: http://people.irisa.fr/Michele.Basseville/kniga/kniga.pdf

[55] Australian Government: Bureau of Meteorology, "ACORN-SAT station adjustment summary - Deniliquin," 2014, Accessed: 11 December 2018. [Online]. Available: http://www.bom.gov.au/climate/change/acorn-sat/documents/station-adjustment-summary-Deniliquin.pdf

[56] Australian Government: Bureau of Meteorology, "ACORN-SAT station adjustment summary - Rutherglen," 2014, Accessed: 11 December 2018. [Online]. Available: http://www.bom.gov.au/climate/change/acorn-sat/documents/station-adjustment-summary-Rutherglen.pdf

[57] M. Begert, T. Schlegel, and W. Kirchhofer, "Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000," *International Journal of Climatology*, vol. 25, no. 1, pp. 65–80, 2005.

[58] J. Reeves, J. Chen, X. L. Wang, R. Lund *et al.*, "A review and comparison of changepoint detection techniques for climate data," *Journal of applied meteorology and climatology*, vol. 46, no. 6, pp. 900–915, 2007.

[59] Australian Government: Bureau of Meteorology, "Long-term temperature record ACORN-SAT: Data and networks," 2018, Accessed: 12 December 2018. [Online]. Available: http://www.bom.gov.au/climate/change/acorn-sat/#tabs=Data-and-networks

[60] Australian Government: Bureau of Meteorology, "Australia's climate data FAQ: (ACORN-SAT)," 2015, Accessed: 12 December 2018. [Online]. Available: http://www.bom.gov.au/climate/change/acorn-sat/documents/ACORN-SAT_TAF_FAQ.pdf

[61] Bureau of Meteorology, "Australian Climate Observations Reference Network - Surface Air Temperature (ACORN-SAT): Station catalogue," Australian Government, Tech. Rep., 2012, Accessed 12 December 2018. [On-

line]. Available: http://www.bom.gov.au/climate/change/acorn-sat/documents/ACORN-SAT-Station-Catalogue-2012-WEB.pdf

[62] M. J. Menne and C. N. Williams Jr, "Homogenization of temperature series via pairwise comparisons," *Journal of Climate*, vol. 22, no. 7, pp. 1700–1717, 2009.

[63] Australian Government: Bureau of Meteorology, "Long-term temperature record ACORN-SAT: Data and networks: Rutherglen Min data," 2019, Accessed: 16 April 2019. [Online]. Available: www.bom.gov.au/climate/change/hqsites/data/temp/tmin.082039.daily.csv

[64] Australian Government: Bureau of Meteorology, "Long-term temperature record ACORN-SAT: Data and networks: Rutherglen Max data," 2019, Accessed: 16 April 2019. [Online]. Available: www.bom.gov.au/climate/change/hqsites/data/temp/tmax.082039.daily.csv

[65] Australian Government: Bureau of Meteorology, "Long-term temperature record: ACORN-SAT Methods," Accessed: 14 December 2018. [Online]. Available: http://www.bom.gov.au/climate/change/acorn-sat/#tabs=Methods

[66] S. L. Barnes, "A technique for maximizing details in numerical weather map analysis," *Journal of Applied Meteorology*, vol. 3, no. 4, pp. 396–409, 1964.

[67] Australian Government: Bureau of Meteorology, "Climate change and variability: Tracker: Australian Time series graphs," 2018, Accessed: 17 December 2018. [Online]. Available: http://www.bom.gov.au/climate/change/

[68] S. J. Torok and N. Nicholls, "A historical annual temperature dataset for Australia," *Australian Meteorological Magazine*, vol. 45, pp. 251–260, 1996.

[69] P. Della-Marta, D. Collins, and K. Braganza, "Updating Australia's high-quality annual temperature dataset," *Australian Meteorological Magazine*, vol. 53, no. 2, pp. 75–93, 2004.

[70] B. C. Trewin, "Extreme temperature events in Australia," Ph.D. dissertation, University of Melbourne, 2001.

[71] B. Trewin, "The Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) version 2," Australian Government: Bureau of Meteorology, Tech. Rep. Bureau Research Report - 032, October 2018, Accessed 16 April 2019. [Online]. Available: http://www.bom.gov.au/research/publications/researchreports/BRR-032.pdf

[72] Australian Government: Bureau of Meteorology, "Long-term temperature record: ACORN-SAT," 2019, Accessed: 16 April 2019. [Online]. Available: http://www.bom.gov.au/climate/data/acorn-sat/

[73] ClimateLab, "ACORN-SAT comparison," 2019, Accessed: 16 April 2019. [Online]. Available: http://climatelab.com.au/acorn/

[74] D. O'Sullivan and D. J. Unwin, *Geographic Information Analysis*, 2nd ed. Hoboken, New Jersey, USA: John Wiley & Sons, 2014.

[75] F. C. Collins, "A comparison of spatial interpolation techniques in temperature estimation," Ph.D. dissertation, Faculty of the Virginia Polytechnic Institute and State University, 1995.

[76] J. Li and A. D. Heap, "A Review of Spatial Interpolation Methods for Environmental Scientists," *Geoscience Australia (Australian Government)*, vol. Geocat #68229, 2008.

[77] W. R. Tobler, "A computer movie simulating urban growth in the Detroit region," *Economic geography*, vol. 46, no. sup1, pp. 234–240, 1970.

[78] R. K. Pace, R. Barry, and C. F. Sirmans, "Spatial statistics and real estate," *The Journal of Real Estate Finance and Economics*, vol. 17, no. 1, pp. 5–13, 1998.

[79] R. Rohde, R. Muller, R. Jacobsen, S. Perlmutter *et al.*, "Berkeley Earth Temperature Averaging Process," *Geoinformatics and Geostatistics: An Overview*, vol. 1, pp. 1–13, 2013.

[80] H. Wackernagel, *Multivariate Geostatistics: An Introduction with Applications*, 3rd ed. Berlin Germany: Springer-Verlag, 2003.

[81] D. G. Krige, "A statistical approach to some basic mine valuation problems on the witwatersrand," *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 52, no. 6, pp. 119–139, 1951.

[82] M. Oliver and R. Webster, "A tutorial guide to geostatistics: Computing and modelling variograms and kriging," *Catena*, vol. 113, pp. 56–69, 2014.

[83] I. Clark, *Practical geostatistics*. Applied Science Publishers London, 2001, vol. 3.

[84] P. A. Burrough, R. McDonnell, and C. D. Lloyd, *Principles of geographical information systems*, 3rd ed. Oxford University Press, 2015.

[85] A. Ghosh, "Efficient thin plate spline interpolation and its application to adaptive optics," Master's thesis, Johannes Kepler University, Linz, Austria, 2010, Available at https://pdfs.semanticscholar.org/cf02/2928311d03aafe71fb7ec6b20885080a72c2.pdf, Accessed 20 December 2018.

[86] R. Daley, *Atmospheric Data Analysis (Cambridge Atmospheric and Space Science Series)*. Cambridge University Press, 1996.

[87] J. K. Eischeid, P. A. Pasteris, H. F. Diaz, M. S. Plantico *et al.*, "Creating a serially complete, national daily time series of temperature and precipitation for the western United States," *Journal of Applied Meteorology*, vol. 39, no. 9, pp. 1580–1591, 2000.

[88] T. Appelhans, E. Mwangomo, D. R. Hardy, A. Hemp *et al.*, "Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania," *Spatial Statistics*, vol. 14, pp. 91–113, 2015.

[89] J. P. Rigol, C. H. Jarvis, and N. Stuart, "Artificial neural networks as a tool for spatial interpolation," *International Journal of Geographical Information Science*, vol. 15, no. 4, pp. 323–343, 2001.

[90] Australian Government: Bureau of Meteorology, "Quality control of climate data," Accessed: 4 January 2019. [Online]. Available: http://www.bom.gov.au/climate/data-services/content/quality-control.html

[91] Igor Zahumensky, "Guidelines on Quality Control Procedures for Data from Automatic Weather Stations," World Meteorological Organization, 2018, Accessed: 30 October 2018. [Online]. Available: https://www.wmo.int/pages/prog/www/OSY/Meetings/ET-AWS3/Doc4(1).pdf

[92] J. J. Marohasy and J. W. Abbot, "Southeast Australian Maximum Temperature Trends, 1887–2013: An Evidence-Based Reappraisal," in *Evidence-Based Climate Science*, 2nd ed., D. J. Easterbrook, Ed. Elsevier, 2016, ch. 5, pp. 83–99.

[93] D. C. Montgomery, *Introduction to statistical quality control*, 6th ed. John Wiley & Sons (New York), 2009.

[94] C. J. Wild and G. A. Seber, "Control charts," in *Chance Encounters: A First Course in Data Analysis and Inference*, A. Battle, Ed. Wiley, 1999, ch. 13.

[95] F. E. James, *Statistical Methods in Experimental Physics*, 2nd ed. Singapore: World Scientific Publishing, 2006.

[96] G. Van Brummelen, *Heavenly Mathematics: The Forgotten Art of Spherical Trigonometry*. Princeton University Press, 2012.

[97] R. L. Burden and J. D. Faires, *Numerical analysis*, 9th ed. Brooks/Cole, 2011.

[98] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[99] S. Deutsch and A. Deutsch, *Understanding the nervous system: an engineering perspective*. IEEE Press, 1993.

[100] F. A. Azevedo, L. R. Carvalho, L. T. Grinberg, J. M. Farfel *et al.*, "Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain," *Journal of Comparative Neurology*, vol. 513, no. 5, pp. 532–541, 2009.

[101] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. D. Jesús, *Neural network design*, 2nd ed. Martin Hagan, 2014, Available at http://hagan.okstate.edu/NNDesign.pdf, Accessed 8 November 2017.

[102] Y. LeCun, L. Bottou, G. Orr, and K.-R. Muller, "Efficient BackProp," in *Neural networks: Tricks of the trade*. Springer, 1998, pp. 9–48.

[103] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014, Accessed: 6 March 2019. [Online]. Available: https://arxiv.org/pdf/1412.6980.pdf

[104] James D. McCaffrey, "Implementing Neural Network L2 Regularization," June 2017, Accessed: 12 March 2019. [Online]. Available: https://jamesmccaffrey.wordpress.com/2017/06/29/implementing-neural-network-l2-regularization/

[105] TensorFlow (Open source machine learning platform), 2019, Accessed: 7 March 2019. [Online]. Available: www.tensorflow.org

[106] GSL - GNU Scientific Library. Accessed: 17 April 2019. [Online]. Available: http://www.gnu.org/software/gsl

[107] Australian Government: Bureau of Meteorology, "Climate Data Online," 2018, Accessed: 9 February 2018. [Online]. Available: http://www.bom.gov.au/climate/data/

[108] Australian Government: Bureau of Meteorology, "Weather Station Directory," 2019, Accessed: 12 April 2019. [Online]. Available: http://www.bom.gov.au/climate/data/stations/

[109] Australian Government: Bureau of Meteorology, "Basic Climatological Station Metadata: Deniliquin (Visitor Information Centre)," 2019, Accessed: 12 April 2019. [Online]. Available: http://www.bom.gov.au/clim_data/cdio/metadata/pdf/siteinfo/IDCJMD0040.074128.SiteInfo.pdf

[110] Matt Mazur, "A step by step backpropagation example," 2019, Accessed: 7 March 2019. [Online]. Available: https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/